



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 138057

**TO:** James Schultz  
**Location:** rem/2d18/2c18  
**Art Unit:** 1635  
**Monday, November 22, 2004**

**Case Serial Number:** 10/001851

**From:** Alex Waclawiw  
**Location:** Biotech-Chem Library  
Rem 1A71  
**Phone:** 272-2534

**[Alexandra.waclawiw@uspto.gov](mailto:Alexandra.waclawiw@uspto.gov)**

### Search Notes

This Page Blank (uspto)

From: Schultz, James  
 Sent: Wednesday, November 17, 2004 8:52 AM  
 To: STIC-Biotech/ChemLib  
 Subject: Seq Search 10/001,851

Hello,  
 Please run a standard amino acid search on SEQ ID NO: 2 from the above entitled application. Instead of truncating the results at the top 15, could you please return the top 50 results from each database as well?  
 Thanks,  
 Doug Schultz

James Douglas Schultz, PhD  
 AU 1635 (Biotechnology)  
 Patent Examiner  
 United States Patent and Trademark Office  
 (Office) REM 2D18  
 (Mail) REM 2C18  
 (571) 272-0763

RECEIVED  
 NOV 17 2004  
 TECH/CHEM/BIOTECH  
 (STIC)

Point of Contact:  
 Alexandra Waclawiw  
 \*\*\*\*\*  
 STAFF USE ONLY Technical Info. Specialist  
 QM1 6A02 Tel 300-449

Searcher: \_\_\_\_\_  
 Searcher Phone: 2-  
 Date Searcher Picked up: 11-22  
 Date Completed: \_\_\_\_\_  
 Searcher Prep/Rev. Time: 11-22  
 Online Time: \_\_\_\_\_

\*\*\*\*\*  
 Type of Search  
 NA Sequence: # 1  
 AA Sequence : #  
 Structure: #  
 Bibliographic: \_\_\_\_\_  
 Litigation: \_\_\_\_\_  
 Patent Family: \_\_\_\_\_  
 Other: \_\_\_\_\_

\*\*\*\*\*  
 Vendors and cost where applicable  
 STN: \_\_\_\_\_  
 DIALOG: \_\_\_\_\_  
 QUESTEL/ORBIT: \_\_\_\_\_  
 LEXIS/NEXIS: \_\_\_\_\_  
 SEQUENCE SYSTEM: \_\_\_\_\_  
 WWW/Internet: \_\_\_\_\_  
 Other(Specify): \_\_\_\_\_

This Page Blank (uspto)

GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

## Om protein - protein search, using sw model

Run on: November 22, 2004, 13:36:41 ; Search time 68 Seconds  
 (without alignments)  
 3181.085 Million cell updates/sec

Title: US-10-001-851-2  
 Perfect score: 328  
 Sequence: 1 MRRKKRLLQAVALVALAALV.....TQQWLFEHTNSTVLRKFFNRN 603

Scoring table: BIOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues  
 Total number of hits satisfying chosen parameters: 2002273  
 Minimum DB seq length: 0  
 Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
 Listing First 100 summaries

Database : A\_Geneseq\_23Sep04:\*

1: geneseqp1980:\*

2: geneseqp1990:\*

3: geneseqp2005:\*

4: geneseqp2005b:\*

5: geneseqp2002b:\*

6: geneseqp2003b:\*

7: geneseqp2004b:\*

8: geneseqp2004b:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match length	DB	ID	Description
1	3278	100.0	603	4	AAU07778	Aau07778 Human nov
2	3278	100.0	603	5	AAU99098	Aau99098 Human gly
3	3272	99.8	603	7	ADC35387	Adc35387 Human N-a
4	3242.5	98.9	631	4	AAU07777	Aau07777 Human nov
5	2771	84.5	506	4	AAU07772	Aau07772 Human nov
6	2771	84.5	506	6	AGT9781	Agt9781 Carbohydr
7	2746.5	83.8	535	4	AAU07771	Aau07771 Human nov
8	1893	57.7	339	4	ABP94733	Abp94733 Human pro
9	1877	57.3	365	4	AAU07776	Aau07776 Human nov
10	1818	55.5	407	4	ARG62600	Aag62600 Human ace
11	1403.5	42.8	319	4	ABG0187	Abg0187 Novel hum
12	1396	42.6	275	7	ADP5865	Adp5865 Human N-a
13	1292	42.5	276	8	ADQ17613	Adq17613 Human nov
14	1385.5	42.3	335	4	ABG0189	Abg0189 Novel hum
15	1155	41.8	269	4	AAU07770	Aau07770 Human nov
16	1360	41.5	666	4	ABB68873	Abb68873 Drosophili
17	1360	41.5	666	4	ABB8867	Abb8867 Drosophili
18	1311	40.0	321	4	AAU07774	Adp07774 Human nov
19	1186	36.2	1252	4	ABP5363	Abp5363 Drosophili
20	1155	35.4	499	4	ABG0190	Abg0190 Novel hum
21	1125	34.3	559	7	ADP5304	Adp5304 Human pol
22	1125	34.3	561	3	ABP43561	Abp43561 Human can
23	1116	34.0	559	7	ADB79836	Adb79836 Rat polyp
24	1116	34.0	559	7	ADB79796	Adb79796 Rat polyp
25	1116	34.0	559	7	ADE62526	Ad62526 Rat Prote

Score	Match length	DB	ID	Description
26	1115	34.0	517	AAR66402
27	1115	34.0	517	AAM16489
28	1115	34.0	559	AAR66397
29	1115	34.0	559	AAR66401
30	1115	34.0	559	AAM16484
31	1113.5	34.0	630	ADP49427
32	1112	33.9	571	ADC10116
33	1106	33.7	556	ABG32379
34	1106	33.7	556	ADE25918
35	1106	33.7	557	ADB76895
36	1096.5	33.5	555	ADC10114
37	1085.5	33.1	573	ADP49427
38	1050	32.0	657	ABP88419
39	1048	32.0	478	ABR1518
40	1048	31.3	478	ADL22696
41	1045	31.9	610	ADP55072
42	1040	31.7	930	AAM18711
43	1035.5	31.6	591	ADP55070
44	1026	31.3	578	ABU07527
45	1024.5	31.3	581	ABG32510
46	1015	30.9	581	ADP55074
47	1013	30.9	940	ADY53040
48	1013	30.9	558	ADK70539
49	999.5	30.5	558	ABP9229
50	999.5	30.5	558	ABP69229
51	991.5	30.2	581	ADC25021
52	991.5	30.2	581	ABE25019
53	991.5	30.2	581	ADH48838
54	990	30.2	240	AAU07773
55	988	30.1	209	AAU07775
56	986.5	30.1	581	ADC5388
57	986.5	30.1	581	ADP00807
58	977.5	29.8	560	ADP25922
59	971.5	29.6	581	ADP55072
60	971	29.6	552	AAK06620
61	971	29.5	552	AAU12269
62	971	29.6	552	AAU09026
63	971	29.6	552	ABO17713
64	971	29.6	552	ABP58585
65	971	29.6	552	ABU0967
66	971	29.6	552	ABU66667
67	971	29.6	552	ABU59748
68	971	29.6	552	ABO24938
69	971	29.6	552	ABU6943
70	971	29.6	552	ADU6146
71	971	29.6	552	ADU18796
72	971	29.6	552	ADU61419
73	971	29.6	552	ADU61420
74	971	29.6	552	ADU77745
75	971	29.6	552	ADU86224
76	971	29.6	552	ADU86224
77	971	29.6	552	ADU86224
78	971	29.6	552	ADU86224
79	971	29.6	552	ADU86224
80	971	29.6	552	ADU86224
81	971	29.6	552	ADU86224
82	971	29.6	552	ADU86224
83	971	29.6	552	ADU86224
84	971	29.6	552	ADU86224
85	971	29.6	552	ADU86224
86	971	29.6	552	ADU86224
87	971	29.6	552	ADU86224
88	971	29.6	552	ADU86224
89	971	29.6	552	ADU86224
90	971	29.6	552	ADU86224
91	971	29.6	552	ADU86224
92	971	29.6	552	ADU86224
93	971	29.6	552	ADU86224
94	971	29.6	552	ADU86224
95	971	29.6	552	ADU86224
96	971	29.6	552	ADU86224
97	971	29.6	552	ADU86224
98	971	29.6	552	ADU86224

Score	Match length	DB	ID	Description
26	1115	34.0	517	AAR66402
27	1115	34.0	517	AAM16489
28	1115	34.0	559	AAR66397
29	1115	34.0	559	AAR66401
30	1115	34.0	559	AAM16484
31	1113.5	34.0	630	ADP49427
32	1112	33.9	571	ADU10116
33	1106	33.7	556	ABG32379
34	1106	33.7	556	ADE25918
35	1106	33.7	557	ADB76895
36	1096.5	33.5	555	ADC10114
37	1085.5	33.1	573	ADP49427
38	1050	32.0	657	ABP88419
39	1048	32.0	478	ABR1518
40	1048	31.3	478	ADU22696
41	1045	31.9	610	ADP55072
42	1040	31.7	930	AAM18711
43	1035.5	31.6	591	ADP55070
44	1026	31.3	578	ABU07527
45	1024.5	31.3	581	ABG32510
46	1015	30.9	940	ADY53040
47	1013	30.9	558	ADY53040
48	1013	30.9	558	ADY53040
49	999.5	30.5	558	ADY53040
50	999.5	30.5	558	ADY53040
51	991.5	30.2	581	ADY53040
52	991.5	30.2	581	ADY53040
53	991.5	30.2	581	ADY53040
54	991.5	30.2	581	ADY53040
55	991.5	30.2	581	ADY53040
56	991.5	30.2	581	ADY53040
57	991.5	30.2	581	ADY53040
58	991.5	30.2	581	ADY53040
59	991.5	30.2	581	ADY53040
60	991.5	30.2	581	ADY53040
61	991.5	30.2	581	ADY53040
62	991.5	30.2	581	ADY53040
63	991.5	30.2	581	ADY53040
64	991.5	30.2	581	ADY53040
65	991.5	30.2	581	ADY53040
66	991.5	30.2	581	ADY53040
67	991.5	30.2	581	ADY53040
68	991.5	30.2	581	ADY53040
69	991.5	30.2	581	ADY53040
70	991.5	30.2	581	ADY53040
71	991.5	30.2	581	ADY53040
72	991.5	30.2	581	ADY53040
73	991.5	30.2	581	ADY53040
74	991.5	30.2	581	ADY53040
75	991.5	30.2	581	ADY53040
76	991.5	30.2	581	ADY53040
77	991.5	30.2	581	ADY53040
78	991.5	30.2	581	ADY53040
79	991.5	30.2	581	ADY53040
80	991.5	30.2	581	ADY53040
81	991.5	30.2	581	ADY53040
82	991.5	30.2	581	ADY53040
83	991.5	30.2	581	ADY53040
84	991.5	30.2	581	ADY53040
85	991.5	30.2	581	ADY53040
86	991.5	30.2	581	ADY53040
87	991.5	30.2	581	ADY53040
88	991.5	30.2	581	ADY53040
89	991.5	30.2	581	ADY53040
90	991.5	30.2	581	ADY53040
91	991.5	30.2	581	ADY53040
92	991.5	30.2	581	ADY53040
93	991.5	30.2	581	ADY53040
94	991.5	30.2	581	ADY53040
95	991.5	30.2	581	ADY53040
96	991.5	30.2	581	ADY53040
97	991.5	30.2	581	ADY53040
98	991.5	30.2	581	ADY53040

Score	Match length	DB	ID	Description
26	1115	34.0	517	AAR66402
27	1115	34.0	517	AAM16489
28	1115	34.0	559	AAR66397
29	1115	34.0	559	AAR66401
30	1115	34.0	559	AAM16484
31	1113.5	34.0	630	ADP49427
32	1112	33.9	571	ADU10116
33	1106	33.7	556	ABG32379
34	1106	33.7	556	ADE25918
35	1106	33.7	557	ADB76895
36	1096.5	33.5	555	ADC10114
37	1085.5	33.1	573	ADP49427
38	1050	32.0	657	ABR88419
39	1048	32.0	478	ABR1518
40	1048	31.3	478	ADU22696
41	1045	31.9	610	ADP55072
42	1040	31.7	930	AAM18711
43	1035.5	31.6	591	ADP55070



XX  
PS Claim 27; FIG 1; 153pp; English.  
XX  
CC The invention relates to the use of a modulator of the activity of an  
CC enzyme such as 47169 or 33935 protein (a novel glycooyl transferase  
CC protein) for making a medicament for modulating the ability of a cell to  
CC affect the glycosylation state of a lipid target or polypeptide target in  
CC a cell. Also included is a method of assessing (M1) if a test compound is  
CC useful for modulating at least one phenomenon (P) such as non-covalent  
CC binding between a protein and one of a cell, a virus and another protein;  
CC cell signalling, cell differentiation, tumourigenesis, cell adhesion, cell  
CC motility, cell-to-cell interaction, cell invasivity, cell proliferation, cell  
CC gene transcription, and an immune response, comprising: (a) adding the  
CC test compound to a first composition comprising a 603 residue 47169  
CC polypeptide sequence (S2), or a sequence at least 30 % identical to a 492  
CC residue 33935 polypeptide sequence (S12), both given in the  
CC ; and (b) comparing the activity in the first composition and in a second  
CC composition that is substantially identical to the first composition,  
CC except that it lacks the test compound, whereby a difference in the  
CC test activity in the first and second compositions is an indication that the  
CC test compound is useful for modulating the phenomenon. The method is  
CC useful for making a medicament for modulating the ability of a cell (e.g.  
CC human endothelial cell such as lung cell, breast cell or colon cell, or  
CC preferably a tumour cell) to affect the glycosylation state of a target  
CC such as lipid or polypeptide. The identified test compound is useful for  
CC treating disorders such as diabetes mellitus, hypoglycaemia, arthritis,  
CC rheumatism, autoimmune disorders, (e.g. systemic lupus erythematosus,  
CC Grave's disease, myasthenia gravis, insulin resistance, rheumatoid  
CC arthritis, scleroderma and autoimmune infertility), tumourigenesis,  
CC cancer and tumour metastasis. The present sequence represents glycosyl  
CC transferase 47169  
XX  
SQ Sequence 603 AA;

Query Match 100.0%; Score 3278; DB 5; Length 603;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;  
Matches 603; Conservative 0;

Qy 1 MRRKEKRLLQAVVALVLAALVLPNGWMLYERQDPTPGGSGAAVAPAGCOSHRSQK 60  
Db 1 MRRKEKRLLQAVVALVLAALVLPNGWMLYERQDPTPGGSGAAVAPAGCOSHRSQK 60

Qy 61 KTFPLGDGOKLKDWDKEAIRRDAQRVNGEQQGPPYPTDAERVDAQYRENGNIVSDK 120  
Db 61 KTFPLGDGOKLKDWDKEAIRRDAQRVNGEQQGPPYPTDAERVDAQYRENGNIVSDK 120

Qy 121 ISLURSLPDPIRHPNCNSKRYLETLPTNTSIIPIFNEGNSSLRITVHSVANRSPPELVARI 180  
Db 121 ISLURSLPDPIRHPNCNSKRYLETLPTNTSIIPIFNEGNSSLRITVHSVANRSPPELVARI 180

Qy 181 VLVQDPSIREHKLPLEDYLALPSVRLRTKREGRLTRMIGASVATGDPVITFLOSH 240  
Db 181 VLVQDPSIREHKLPLEDYLALPSVRLRTKREGRLTRMIGASVATGDPVITFLOSH 240

Qy 241 BANVNLWPLDRTARNRKTIVCMIDVDDDRYETQAGDAMRGAFDWMYKRIPIP 300  
Db 241 BANVNLWPLDRTARNRKTIVCMIDVDDDRYETQAGDAMRGAFDWMYKRIPIP 300

Qy 301 PELQKADRSDFPESPVAGLFLAVDRKWNWLSLGSLSLIGSCEBYESPCKYNGERN 360  
Db 301 PELQKADRSDFPESPVAGLFLAVDRKWNWLSLGSLSLIGSCEBYESPCKYNGERN 360

Qy 361 EDIPCSRVKHYIYRCKVYPUVPAVSLARNLKVRAEYMDDEAYTQORPEYRHLASDV 420  
Db 361 EDIPCSRVKHYIYRCKVYPUVPAVSLARNLKVRAEYMDDEAYTQORPEYRHLASDV 420

Qy 421 AVQKQLRSLSNCSSPKWEMTKIANDLPKTPYPTPEPAWGEIRNGLQGCAUTKHALG 480  
Db 421 AVQKQLRSLSNCSSPKWEMTKIANDLPKTPYPTPEPAWGEIRNGLQGCAUTKHALG 480

Qy 481 SPURLEGCVRGZAAWNMNOVFTWDRDIPDQHCKFCDAISHTSPVLYDCHS 540  
Db 481 SPURLEGCVRGZAAWNMNOVFTWDRDIPDQHCKFCDAISHTSPVLYDCHS 540

Qy 541 MKGNOLWKRKDKKLYHVGSCCMDCSSEDRIFMNTCNPSSLTQOMFREHTNSTVLPK 600  
Db 541 MKGNOLWKRKDKKLYHVGSCCMDCSSEDRIFMNTCNPSSLTQOMFREHTNSTVLPK 600  
Qy 601 NRN 603  
Db 601 NRN 603

RESULT 3  
ID ADC35387  
ID ADC35387 standard; protein; 603 AA.  
XX  
XX  
AC ADC35387;  
XX  
DT 18-DEC-2003 (first entry)  
DB Human N-acetylgalactosamine transferase (GalNAc) T13 protein.  
XX  
KW N-acetylgalactosamine transferase; GalNAc; alpha1-bond;  
KW genetic engineering; GalNAc-T13; cytostatic; cancer treatment;  
KW cancer diagnosis; gene therapy; human; enzyme.  
XX  
OS Homo sapiens.  
XX  
WO2003057887-A1.  
XX  
PD 17-JUL-2003.  
XX  
PF 06-JAN-2003; 2003WO-JP000008.  
XX  
PR 28-DEC-2001; 2001JP-00401507.  
PR 05-JUN-2002; 2002JP-00163822.  
PR 12-JUL-2002; 2002JP-00203626.  
XX  
PA (NAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
PA (AMSH) AMERSHAM BIOSCIENCES KK.  
PA (FURE) FUJIREBIO INC.  
XX  
PI Narimatsu H, Zhang Y, Gotoh M;  
DR WPI; 2003-587133/55.  
XX  
N-PSDB; ADC35394.

PT Novel UDP-N-acetyl- $\beta$ -D-galactosamine:polypeptide N-acetylgalactosamine  
PT transferase and encoded nucleic acid, applicable in identifying O-bonded  
PT sugar-attached proteins, gene diagnosis, and cancer therapy.  
XX  
PS Claim 1; SEQ ID NO 3; 153pp; Japanese.

XX  
CC This invention relates to a novel protein with an activity of  
CC transferring N-acetylgalactosamine (GalNAc) via an alpha1-bond to a  
CC hydroxyl group of serine or threonine in a protein or a peptide sequence.  
CC The invention provides gene and protein sequences for these enzymes to  
CC enable genetic engineering or production of these enzymes. Seven genes  
CC and their proteins are described by the invention, GalNAc-T1 to 17. The  
CC enzymes of the invention may have cytostatic activity, and hence may be  
CC used for treatment or diagnosis of cancer, and the sequences described  
CC may be useful in gene therapy. The present sequence is the partial  
CC sequence of the human GalNAc-T13 transferase protein of the invention.  
XX  
SQ Sequence 603 AA;

Query Match 99.8%; Score 3272; DB 7; Length 603;  
Best Local Similarity 99.8%; Pred. No. 0; Mismatches 1; Indels 0; Gaps 0;  
Matches 602; Conservative 1;

Qy 1 MRRKEKRLLQAVVALVLAALVLPNGWMLYERQDPTPGGSGAAVAPAGCOSHRSQK 60  
Db 1 MRRKEKRLLQAVVALVLAALVLPNGWMLYERQDPTPGGSGAAVAPAGCOSHRSQK 60



RESULT 5	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
AAU07772	ID	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	XX	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
AC	XX	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
DT	DT	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
KW	OS	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
Human; transferase; breast cancer; prostate cancer; immunogen;	Homo sapiens.	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	PN	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	PD	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	PF	28-FEB-2001;	2001WO-US006460.	
XX	PR	29-FEB-2000;	2000US-0186558P.	
XX	PR	02-MAR-2000;	2000US-0191849P.	
XX	PA	(LEXI-) LEXICON GENETICS INC.		
XX	PI	Donoho G, Hilbun B, Turner CA, Friedrich G, Abuin A, Kieke JA;		
XX	PI	Zambrowicz B, Sands AT, Walko DW, Wilganowski NL, Hu Y, Potter DG;		
XX	DR	WPI; 2001-550185/61.		
XX	DR	N-PSDB; AAS12619.		
XX	PT	Novel nucleic acid sequences encoding novel human proteins useful for diagnosis, drug screening, clinical trial monitoring and treatment of diseases and disorders.		
XX	PT	Novel nucleic acid sequences encoding novel human proteins useful for diagnosis, drug screening, clinical trial monitoring and treatment of diseases and disorders e.g. breast and prostate cancer. NHPs can also be useful for augmenting the efficacy of chemotherapeutic agents used in treatment of breast or prostate cancer. The nucleic acid is also useful in NHP gene regulation, and as antisense primers in amplification reactions of NHP gene sequences. NHPs are useful for producing antibodies. The present sequence represents a novel human transferase which has sequence similarity to N-acetyl-galactosaminyltransferase		
XX	PS	Claim 8; Page 47-48; 60pp; English.		
XX	CC	The invention relates to isolated nucleic acids encoding novel human transferase proteins (NHP). The nucleic acids and proteins are useful for diagnosis, drug screening, clinical trial monitoring and treatment of diseases and disorders e.g. breast and prostate cancer. NHPs can also be useful for augmenting the efficacy of chemotherapeutic agents used in treatment of breast or prostate cancer. The nucleic acid is also useful in NHP gene regulation, and as antisense primers in amplification		
CC	CC	reactions of NHP gene sequences. NHPs are useful for producing antibodies. The present sequence represents a novel human transferase which has sequence similarity to N-acetyl-galactosaminyltransferase		
XX	CC	Sequence 506 AA;		
XX	SQ	Query Match		
Best Local Similarity	84.5%	Score	2771; DB 4; Length	506;
Matches	506;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	98	MTDAERDVAQRENGPNIVDKISLANSLDPDIRHPCNSKRIELTNTSIIPTFHNG	157	
Db	1	MTDAERDVAQRENGPNIVDKISLANSLDPDIRHPCNSKRIELTNTSIIPTFHNG	60	
QY	158	WSSLRLRTWHSVLRNSPPELVAEVTLVDDFSDRSHLKKPLEDWALFSPVRIKTKRGL	217	
Db	61	WSSLRLRTWHSVLRNSPPELVAEVTLVDDFSDRSHLKKPLEDWALFSPVRIKTKRGL	120	
QY	218	IRTRMLGASVATGIVTELDSHRERANVWLPPLDRLARNRKTIVCPMDVDDHDFYE	277	
Db	121	IRTRMLGASVATGIVTELDSHRERANVWLPPLDRLARNRKTIVCPMDVDDHDFYE	180	
QY	278	TQGDAMGAFDMMYKRPITPELQADPSPPESVOMGGLPAVDRKWFELLGDP	337	
Db	181	TQGDAMGAFDMMYKRPITPELQADPSPPESVOMGGLPAVDRKWFELLGDP	240	
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	DB	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	QY	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	360
XX	QY	458	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	517
XX	DB	361	AANGEIRVNGTGLCAGTGHAGSPRLGCVGRGELAANNQVPTTWRERIRPGPQ	420
XX	QY	518	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	577
XX	DB	421	HTKKEFCFDASHTSPVTLYDCHENKGNDLWKYRKDCTLYHPVSGSCMDCSESCHRIFTMT	480
XX	QY	578	CNSSLTQWLFEHTNSTVLEKRN	603
XX	DB	481	CNSSLTQWLFEHTNSTVLEKRN	506
XX	QY	338	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	397
XX	AC	241	GLEIWGGBOYEISPKVWMCGGRMEDIPSPRSVGHLYRKVYPUKUPAGVSLARNJKVAEW	300
XX	AC	398	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKTWMKMTKIAWDLPKFYPVPEA	457
XX	DB	301	MDEYAEYTYQRPRYRHSAGDVAVKLRSNCKSKT	

PN	WO200291060-A2.
XX	05-DEC-2002.
PD	
XX	22-MAY-2002; 2002WO-US018354.
PF	
XX	25-MAY-2001; 2001US-0233768P.
PR	01-AUG-2001; 2001US-0309548P.
PR	23-AUG-2001; 2001US-0314400P.
PR	19-OCT-2001; 2001US-0443706P.
PR	07-DEC-2001; 2001US-0337999P.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PT	Swarnakar A, Gorvad AE, Hafalia AJA, Duggan BM, Emerling BM;
PI	Ison CH, Nguyen DB, Lee EA, Yue H, Yue H, Forsythe JU, Li JX;
PT	Thangavelu K, Walia NK, Burford N, Mason PM, Lal PG, Lee S;
PI	Becha SD, Tang YT;
XX	
DR	WPI; 2003-140462/13.
DR	N-PSDB; ABA00833.
XX	
PT	Novel human carbohydrate associated polypeptide, useful in diagnosis, treatment and prevention of carbohydrate metabolism, cell proliferative, autoimmune/inflammatory, reproductive, and neurological disorders.
XX	
PS	Claim 1; Page 126-27; 141PP; English.
XX	
CC	The sequences given in AAG7979-88 represent human carbohydrate associated polypeptides (CHOP). The CHOP polypeptides and the nucleotide sequences encoding them, are useful for diagnosing, treating and preventing cancer, carbohydrate metabolic disorders (e.g. diabetes, anemia, hypoglycemia, obesity, glycogen storage disease, neurological dysfunctions), cell proliferative disorders (e.g. acinic keratosis, arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, psoriasis), autoimmune/inflammatory disorders (such as acquired immunodeficiency syndrome (AIDS), Addison's disease, allergies, asthma, contact dermatitis, bronchitis, gout, Hashimoto's thyroiditis), viral, bacterial, fungal, parasitic, protozoal and helminthic infections, reproductive disorders (e.g. endometriosis), polycystic ovary syndrome, ovarian hyperstimulation syndrome), genetic disorders (e.g. Down syndrome, cystic fibrosis, sickle cell anaemia, thalassemia), cardiac disorders (e.g. myocarditis, cardiomyopathy), neurological disorders (e.g. Alzheimer's disease, Partinon's disease, dementia, depression, epilepsy, Tourette's syndrome (e.g. cerebral palsy), mental disorders (e.g. mood, anxiety). CHOP proteins are useful in a number of drug screening techniques, and to analyse the proteome of a tissue or cell type. CHOP cDNA is useful for creating "knockin" humanized animals or transgenic animals to model human diseases in somatic or germline gene therapy, to generate a transcript image of a tissue or cell type, for detecting differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier or affected individuals, and as hybridization probes for mapping naturally occurring genomic sequences
CC	Sequence 506 AA;
XX	
Query Match	84.5%; Score 2771; DB 6; Length 506;
Best local Similarity	100.0%; Pred. No. 8.2e-274;
Matches	506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	98 MTDAAEVDQWRENGPNIVYSDKLISNRSLPDIRPHNCNSKRYLETLPLWNTSITIPFHNEG
Db	1 MTDAAEVDQWRENGPNIVYSDKLISNRSLPDIRPHNCNSKRYLETLPLWNTSITIPFHNEG
Qy	158 WSSLRTHVSLNRSPPPELVAVIYVDDPDREREHKKPLDYMALPSVRLIRTKREGI
Db	61 WSSLRTHVSLNRSPPPELVAVIYVDDPDREREHKKPLDYMALPSVRLIRTKREGI
Qy	218 ITRMLGASVATGDVTFPLSHCEANVNLPLPPLDRIARAKRTVCPMTDVIDHDFYE
Db	121 ITRMLGASVATGDVTFPLSHCEANVNLPLPPLDRIARAKRTVCPMTDVIDHDFYE
SQ	Sequence 506 AA;
XX	
Qy	278 TQGCDAMGAFDNEYKRIPPELQADPSDPFSPVAGLFAVTRKWFELGGDP 337
Db	181 TQGCDAMGAFDNEYKRIPPELQADPSDPFSPVAGLFAVTRKWFELGGDP 240
Qy	338 GLEWGGGEOYEISPKWVCGGRMEDIPSRVGHYTKVYKUPAGVSLARNKRVAEW 397
Db	241 GLETWGGGEOYEISPKWVCGGRMEDIPSRVGHYTKVYKUPAGVSLARNKRVAEW 300
Qy	398 MDEVAEYVYQRRPBYRHISAGDVAVOKURSSLANCKSTKWFMTKLAWDLPKFVPPVERPA 457
Db	301 MDEVAEYVYQRRPBYRHISAGDVAVOKURSSLANCKSTKWFMTKLAWDLPKFVPPVERPA 360
Qy	458 AAAGCBEIRVNGTGICADTKIGALGSPLRLEGYVGRGEBAWNNOVFTTWREDIRPGPQ 517
Db	361 AAAGCBEIRVNGTGICADTKIGALGSPLRLEGYVGRGEBAWNNOVFTTWREDIRPGPQ 420
Qy	518 HTKURCPDISHSPVTYDCHSMKGQNLWKRKDQYHYPGSCMDCESPHRFNT 577
Db	421 HTKURCPDISHSPVTYDCHSMKGQNLWKRKDQYHYPGSCMDCESPHRFNT 480
Qy	578 CNPASLTQOWLFHNTNSTLKEPNR 603
Db	481 CNPASLTQOWLFHNTNSTLKEPNR 506
RESULT 7	
AAU0771	AAU0771 standard; protein; 535 AA.
XX	
AC	AAU0771;
XX	
DT	04-DEC-2001 (first entry)
XX	
DE	Human novel transferase protein, NHP #14.
XX	
KW	Human; transferase; breast cancer; prostate cancer; immunogen; gene therapy; antisense.
XX	
OS	Homo sapiens.
XX	
PN	WO200161903-A2.
XX	
PD	07-SEP-2001.
XX	
PF	28-FEB-2001; 2001WO-US006460.
XX	
PR	29-FEB-2000; 2000US-0185920P.
PR	02-MAR-2000; 2000US-0186558P.
PR	24-MAR-2000; 2000US-0191849P.
XX	
PA	(LEXI-) LEXICON GENETICS INC.
XX	
PI	Domoh G, Hilbun E, Turner CA, Friedrich G, Abuin A, Wilganowski NL, Hu Y, Kieke JA, Porter DG;
PI	Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kieke JA;
XX	
DR	WPI; 2001-550185/61.
DR	N-PSDB; AAS12618.
XX	
PT	Novel nucleic acid sequences encoding novel human proteins useful for diagnosis, drug screening, clinical trial monitoring and treatment of diseases and disorders.
XX	
PS	Claim 7; Page 45-46; 60PP; English.
XX	
CC	The invention relates to isolated nucleic acids encoding novel human transerase proteins (NHP). The nucleic acids and proteins are useful for diagnosis, drug screening, clinical trial monitoring and treatment of diseases and disorders e.g. breast and prostate cancer. NHPs can also be useful for augmenting the efficacy of chemotherapeutic agents used in treatment of breast or prostate cancer. The nucleic acid is also useful in NHP gene regulation, and as antisense primers in amplification reactions of NHP gene sequences. NHPs are useful for producing

CC antibodies. The present sequence represents a novel human transferase  
 CC which has sequence similarity to N-acetyl-galactosaminyltransferase  
 XX  
 SQ Sequence 535 AA:

Query Match 83.8%; Score 2745.5; DB 4; Length 535;  
 Best Local Similarity 94.6%; Pred. No. 2.9e-271; Mismatches 0; Indels 29; Gaps 1;  
 Matches 506; Conservative 0; Sequence 535 AA:

QY 98 MTDAERDVDAKRENGPNIVSKISLNRLPDIRHPCNSKRVLETLPTNSITIPFHNEG 157  
 1 MTDAERDVDAKRENGPNIVSKISLNRLPDIRHPCNSKRVLETLPTNSITIPFHNEG 60  
 DB 158 WSSLURTVHSVLRSRPPPEVAVTIVLDDSDREHLKKPLDVMALFPSPVRLKTKREGI 217  
 61 WSSLURTVHSVLRSRPPPEVAVTIVLDDSDREHLKKPLDVMALFPSPVRLKTKREGI 120  
 QY 218 IRTMLGASVATGIVTFLDSDCHCANVNLPPDPESPVMAAGIIPAVDRKWFWBLG3DP 277  
 121 IRTMLGASVATGIVTFLDSDCHCANVNLPPDPESPVMAAGIIPAVDRKWFWBLG3DP 180  
 QY 278 TAGDAMRGAPDWMYKIPIPPELQADPSDPESPVMAAGIIPAVDRKWFWBLG3DP 337  
 181 TAGDAMRGAPDWMYKIPIPPELQADPSDPESPVMAAGIIPAVDRKWFWBLG3DP 240  
 DB 338 GLEIWGGEYEISPK----- VWMCGGRMEDIPCSRV 368  
 241 GLEIWGGEYEISPKGLIHLPLRVLNSWFOAVLPRAPMMLQVWMCGGRMEDIPCSRV 300  
 QY 369 GHYRKVYKPYKVPAGVSLARNLKEVAVWMDAEYIVYORRPEYRHLASGDAVOKLRS 428  
 301 GHYRKVYKPYKVPAGVSLARNLKEVAVWMDAEYIVYORRPEYRHLASGDAVOKLRS 360  
 DB 429 SLNCKSKFMFMTKIAWDLKPYPVPEPPAAMWGEIRNVGTGLCADTRKGALGSPRLGC 488  
 361 SLACKSFPEWMTKIAWDLKPYPVPEPPAAMWGEIRNVGTGLCADTRKGALGSPRLGC 420  
 QY 489 VRGEAAMWNNMQVFTFWREDTRPGDROHTKCPFDASHTSPVTLDCHSMGNOLWK 548  
 421 VRGEAAMWNNMQVFTFWREDTRPGDPOHTKCPFDASHTSPVTLDCHSMGNOLWK 480  
 QY 549 YRKDKTLIHPVSCSMDSESDRIFMTNCPSLTOQMLFENTNSTYLEKENRN 603  
 481 YRKDKTLIHPVSCSMDSESDRIFMTNCPSLTOQMLFENTNSTYLEKENRN 535

RESULT 8

AB94733 AB94733 standard; protein; 339 AA.

ID AB94733;

AC AB94733;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:15766.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EF1074617-A2.

XX PD 07-FEB-2001.

XX PP 28-JUL-2000; 2000EP-00116126.

XX PR 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

PA (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PT Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX DR WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX PS Claim 8; SEQ ID NO 15766; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the 5602 nucleotide sequence comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the 5'-end sequence of the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH0166 to AAH13628 and AAH13633 to AAH1742 represent human cDNA sequences; AAH2446 to AAH5933 represent human amino acid sequences; and AAH3623 to AAH3632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX SQ Sequence 339 AA;

Query Match 57.7%; Score 1993; DB 4; Length 339;  
 Best Local Similarity 99.7%; Pred. No. 3.1e-184; Mismatches 0; Indels 0; Gaps 0;  
 Matches 338; Conservative 1; Sequence 339 AA:

QY 265 MIDVIDHDFRYETOQAGDAMRGAPDWMYKIPIPPELQADPSDPESPVMAAGIIPAV 324  
 1 MIDVIDHDFRYETOQAGDAMRGAPDWMYKIPIPPELQADPSDPESPVMAAGIIPAV 60

DB 325 DRKWFWEGLGGDGLTIGGEGEOEISFPVWMCGGRMEDIPCSRVGHYKRYKPYKVPAGV 384  
 61 DRKWFWEGLGGDGLTIGGEGEOEISFPVWMCGGRMEDIPCSRVGHYKRYKPYKVPAGV 120

QY 385 SLARNLKEVAVWMDAEYIVYORRPEYRHLASGDAVOKLRSLLACKSKFMFTKAW 444  
 121 SLARNLKEVAVWMDAEYIVYORRPEYRHLASGDAVOKLRSLLACKSKFMFTKAW 180

DB 445 DLKPYPVPEPPAAMWGEIRNVGTGLCADTRKGALGSPRLGCVRGEGAAAMWNNMQVFT 504  
 101 DLKPYPVPEPPAAMWGEIRNVGTGLCADTRKGALGSPRLGCVRGEGAAAMWNNMQVFT 240

QY 505 PTREDIDTRPGDROHTKCPFDASHTSPVTLDCHSMGNOLWKYRKDKTLIHPVSCM 564  
 241 PTREDIDTRPGDPOHTKCPFDASHTSPVTLDCHSMGNOLWKYRKDKTLIHPVSCM 300

QY 555 DCSESDRIFMTNCPSLTOQMLFENTNSTYLEKENRN 603  
 301 DCSESDRIFMTNCPSLTOQMLFENTNSTYLEKENRN 339

RESULT 9

AAU0776

ID AAU0776 standard; protein; 366 AA.

XX AC AAU0776;

XX PA

DT	04-DEC-2001	(first entry)	Db	301 PELOKADSPDPESPVMAGGLFAVDKRNFWELGGYDPOLEINGGEQYBISPKV
XX	XX	DE Human novel transferase protein, NHP #19.	RESULT 10	AAAG62600
XX	XX	KW Human; transferase; breast cancer; prostate cancer; immunogen;	ID AAAG62600	standard; protein; 407 AA.
XX	XX	KW gene therapy; antisense.	ID	AAAG62600/
OS	XX	Homo sapiens.	AC	
PN	XX	WO200164903-A2.	DT	06-SEP-2001 (first entry)
XX	XX	PR 07-SEP-2001.	XX	Human acetyl galactosyl transferase 45.
XX	XX	PR 28-FEB-2001; 2001WO-US006460.	XX	Human; acetyl galactosyl transferase 45; GalNAc-T45; cancer; haemopathy;
XX	XX	PR 29-FEB-2000; 2000US-0185920P.	XX	HIV infection; immunological disease; inflammation; gene therapy;
PR	22-MAR-2000; 2000US-0186558P.	XX	Homo sapiens.	
PR	24-MAR-2000; 2000US-0193849P.	XX	WO200138545-A1.	
XX	XX	(LEXI-) LEXICON GENETICS INC.	XX	
PA	XX	XX	XX	
PI	XX	Domoto G, Hilbun E, Turner CA, Friedrich G, Abuin A; Kleke JA;	XX	
PI	XX	PI Kamrowicz B, Sands AT, Walko DW, Wilganowski NL, Hu Y, Kleke JA;	XX	
PT	XX	PI Porter DG;	XX	
PT	XX	WPI; 2001-550185/61.	XX	
PT	XX	DR N-PSDB; AAS12623.	XX	
PT	XX	Novel nucleic acid sequences encoding novel human proteins useful for diagnosis, drug screening, clinical trial monitoring and treatment of diseases and disorders.	XX	
PT	XX	PT diagnosis, drug screening, clinical trial monitoring and treatment of diseases and disorders.	XX	
PS	XX	Disclosure; Page 52-53; 60pp; English.	XX	
CC	CC	The invention relates to isolated nucleic acids encoding novel human transferrase proteins (NHP). The nucleic acids and proteins are useful for diagnosis, drug screening, clinical trial monitoring and treatment of diseases and disorders e.g. breast and prostate cancer. NHPs can also be useful for augmenting the efficacy of chemotherapeutic agents used in treatment of breast or prostate cancer. The nucleic acid is also useful in NHP gene regulation, and as antisense primers in amplification reactions of NHP gene sequences. NHPs are useful for producing antibodies. The present sequence represents a novel human transferase which has sequence similarity to N-acetyl-galactosaminyltransferase	XX	
CC	CC	Sequence 366 AA;	XX	
SQ	XX	Query Match 57.3%; Score 1877; DB 4; Length 366; Best Local Similarity 100.0%; Pred. No. 1.5e-182; Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	XX	
QY	1	1 MRRKKERLQLQAVALYALVALYLPNGLWALTYREROPDGTRGGSGAVAPAGGSHSRK 60	QY	200 MALPPSVLRLTRKREGIIRTRMIGASATGDPVITPLSHCEANVNWLPLDRIARRK 259
Db	1	1 MRRKKERLQLQAVALYALVALYLPNGLWALTYREROPDGTRGGSGAVAPAGGSHSRK 60	Db	1 MARFSKVRVTRKREGIIRTRMIGASATGDPVITPLSHCEANVNWLPLDRIARRK 60
QY	61	61 KTEFLGDGQKLDWHDKEARRDAGQVNGEGRGPYPMTDARVQDQYRBNFNYSDK 120	QY	260 TIVCPMDIVDHDPRYEQYQAGDAMRGAFDEMWMYKRPPIPSLQKADPSDPRESPWAG 319
Db	61	61 KTEFLGDGQKLDWHDKEARRDAGQVNGEGRGPYPMTDARVQDQYRBNFNYSDK 120	Db	61 TIVCPMDIVDHDPRYEQYQAGDAMRGAFDEMWMYKRPPIPSLQKADPSDPRESPWAG 319
QY	121	121 ISLNSLPLDPRHPCNCNSKRYLETLNTSIIIPFHNGWSSLRHTHVSUNRSPBLVATI 180	QY	320 GLFAVDRKRNFWELGGYDPSLBNWEGQWISFKWMC3MEDIPCSRRGHTIRKVKYVK 379
Db	121	121 ISLNSLPLDPRHPCNCNSKRYLETLNTSIIIPFHNGWSSLRHTHVSUNRSPBLVATI 180	Db	121 GLFAVDRKRNFWELGGYDPSLBNWEGQWISFKWMC3MEDIPCSRRGHTIRKVKYVK 379
QY	181	181 VLVDDFSDRKKLKCLEDYMLAFPSVRLRKKRGLIRTRMIGASVATGDNVITPLDSIC 240	QY	380 VPAGVSLAARNLKRVAEWMDYBASYTQRPYRPEYHLSGAGDVAVOKKLSSLNQSKPKFM 439
Db	181	181 VLVDDFSDRKKLKCLEDYMLAFPSVRLRKKRGLIRTRMIGASVATGDNVITPLDSIC 240	Db	181 VPSGJSLAARNLKRVAEWMDYBASYTQRPYRPEYHLSGAGDVAVOKKLSSLNQSKPKFM 439
QY	241	241 EANVWHLPLDRIARNKRTIVCPMDIVDHDPRYEQYQAGDAMRGAFDEMWMYKRPIP 300	QY	440 TKIADLPLKRYPPVPPAAGWEIRNWTGLCADTKHAGLSPDLESCVGRGHEAMNN 499
Db	241	241 EANVWHLPLDRIARNKRTIVCPMDIVDHDPRYEQYQAGDAMRGAFDEMWMYKRPIP 300	Db	241 AAAVADPKYPPVPPAAGWEIRNVAALCYTDKSGATGTBLRDLICVKDSEWTSH 300
QY	301	301 PELQKADSPDPESPVMAGGLFAVDKRNFWELGGYDPOLEINGGEQYBISPKV 353	QY	500 MQVFTFTWREDIRPGDPQMTKKFCFDATISHTSPVTLVYDCHSMKQNQLWKRKDKTLYHPV 559

Db 301 EQLPTFGWREDIRPGEPLTRKCPFDATSHNSPVTLYDCHGMKGQNQLQYRKYRKTLEHPV 360  
 Qy 560 SGSCMDCSBDRHIFMNTCNPSSITQOMIPEHTNSTVIEKFRN 603  
 Db 361 SNSOMDCBPAEKIFMACDPLSTQWTFEHTNSTVIEKRNH 404

RESULT 11  
 ID ABG08187  
 ID ABG08187 standard; protein; 319 AA.  
 XX AC ABC08187;  
 XX DT 13-FEB-2002 (first entry)  
 XX DB Novel human diagnostic protein #8178.  
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN WO200175057-A2.  
 XX PD 11-OCT-2001.  
 XX PR 30-MAR-2001; 2001WO-US008631.  
 XX PR 31-MAR-2000; 2000US-00540217.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Dumanac RT, Liu C, Tang YT;  
 XX DR WPI; 2001-639362/73.  
 XX PR New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations PT responsible for genetic disorders or other traits and to assess PT biodiversity.  
 XX PS Claim 20; SEQ ID NO 39546; 103PP; English.  
 XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 319 AA;

Db 302 AGVSLARSKNLRKVAEVNDEAVYIYVORRPEYRHLSGDVAVOKKRSLSACKSPKFM 439  
 Qy 71 AGVSLARSKNLRKVAEVNDEAVYIYVORRPEYRHLSGDVAVOKKRSLSACKSPKFM 130  
 Db 440 TKIAWDLRKFYPPVEPPAAGFIRNGTGLCADTRKGALGSPDRLRGCVRGRGEAANN 499  
 Qy 500 MQVFTFREDIRPGEPLTRKCPFDATSHNSPVTLYDCHGMKGQNQLQYRKYRKTLEHPV 559  
 Db 191 MQVFTFREDIRPGEPLTRKCPFDATSHNSPVTLYDCHGMKGQNQLQYRKYRKTLEHPV 250  
 Qy 560 SGSCMDCSBDRHIFMNTCNPSSITQOMIPEHTNSTVIEKRN 603  
 Db 251 SGSCMDCSBDRHIFMNTCNPSSITQOMIPEHTNSTVIEKRN 294

RESULT 12  
 ID ADE25865  
 ID ADE25865 standard; protein; 275 AA.  
 XX AC ABC25865;  
 XX DT 29-JAN-2004 (first entry)  
 XX PR Human N-acetylgalactosaminyl transferase T10.  
 XX PA Cytostatic; respiratory; antiasthmatic; antiinflammatory; endocrine; CNS; immunosuppressive; antiartritic; antibacterial; vulnerability; GalNAc-transferase; lectin domain inhibitor; N-acetylgalactosaminyl transferase; GalNAc-transferase; GalNAc-T4; GalNAc-T7; GalNAc-T2; GalNAc-T3; GalNAc-transferase; tumour; cancer; lung disease; mucous accumulation; asthma; chronic bronchitis; smokers lung; cystic fibrosis; exocrine gland disease; mucin secretion decrease; Sjogren's syndrome; dry mouth; selectin-mediated leukocyte trafficking; autoimmunity; arthritis; leukaemia; lymphoma; immunosuppression; sepsis; wound healing; inflammation; mucin secretion; hypersecretion; mucin accumulation; chronic obstructive respiratory pulmonary disease; asthma; lectin domain inhibition; GalNAc-glycopeptide; Galbeta1-3GalNAc-glycopeptide; UDP-Gal donor; glycosyltransferase; core 1 beta1,3-galactosyltransferase; alpha2,6-sialyltransferase; N-acetylgalactosaminyl transferase 4Mpa2 glycopeptide; O-glycosylation; breast cancer; human; GalNACT10.  
 XX OS Homo sapiens.  
 XX PN US2003186850-A1.  
 XX PD 02-OCT-2003.  
 XX PR 12-NOV-2002; 2002US-00292896.  
 XX PR 11-MAY-2000; 2000US-0203331P.  
 XX PR 10-MAY-2001; 2001WO-DK000228.  
 XX PR 08-NOV-2002; 2002US-0425204P.  
 XX PA (GLYC-) GLYCOCYIM APS.  
 XX PI Clausen H, Bennett EP, Hassan H, Reis CA;  
 XX DR WPI; 2003-831259/77.

Query Match Similarity 42.8%; Score 1403.5; DB 4; Length 319; Best Local Similarity 90.8%; Pno. 3.5e-134; Matches 258; Conservative 4; Mismatches 19; Indels 3; Gaps 2;

Modulating N-acetylgalactosaminyltransferase functions comprises administration of an N-acetylgalactosaminyltransferase lectin domain inhibitor.

Disclosure; Page 11; 65pp; English.

XX  
 CC The invention describes a method of modulating polypeptide N-  
 CC acetylgalactosaminyl (GalNAc)-transferase functions comprising  
 CC administration of a GalNAc-transferase lectin domain inhibitor (I). (I)  
 CC is used for modulating functions of GalNAc-transferase (e.g. GalNAc-T4,  
 CC GalNAc-T7, GalNAc-T2), for inhibiting one Lectin  
 domain of the GalNAc-transferase and modulating the function mediated by  
 the domain for preparing a medicament for treating tumours and cancers,  
 CC lung diseases associated with mucous accumulation (e.g. asthma, chronic  
 CC bronchitis, smoker's lung, and cystic fibrosis), diseases of exocrine  
 glands associated with increased or decreased mucin secretion (e.g.  
 CC Siegelman's syndrome and dry mouth), disorders associated with  
 CC dysregulation of selectin-mediated leukocyte trafficking (e.g.  
 CC autoimmunity, arthritis, leukaemias, lymphomas, immunosuppression,  
 CC sepsis, wound healing, acute and chronic inflammation); inhibiting mucin  
 CC secretion; and inhibiting hypersecretion and accumulation of mucin in the  
 lungs of mammals (e.g. humans) suffering from chronic obstructive  
 CC respiratory pulmonary disease, asthma, and cystic fibrosis. (I) is  
 effective in inhibiting at least one lectin domain of the GalNAc-  
 CC transferase and modulating the inhibited function mediated by the domain  
 (preferably GalNAc-glycopeptide, Galbeta1-GalNAc-glycopeptide, or  
 peptide specificity of the transferase, or UDP-Gal donor substrate  
 specificity). (I) Selectively inhibits at least one member of the GalNAc-  
 transferase family without inhibiting other glycosyltransferases  
 (preferably core 1 beta1,3-galactosyltransferase, alpha2,6-  
 sialyltransferase, or glycosidase functioning in the O-  
 CC glycosylation pathway). This is the amino acid sequence of human N-  
 CC acetylgalactosaminyl transferase T10.  
 XX

SEQ ID: 275 AA;

Query Match 42.6%; Score 1396; DB 7; Length 275;  
 Best Local Similarity 93.7%; Pred. No. 1.6e-133;  
 Matches 233; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 336 DPGLSITWGSGQQEYLSFKVNRGCGGRMEDIPCSRSVGHYIYRKVYVVKYVPAVGSLA--RNLRV 393  
 6 DGELEBTSSFLPAMQWNGCGGRMEDIPCSRSVGHYIYRKVYVVKYVPAVGSLA--RNLRV 393  
 394 AEWVNDAYAIXYIYORPEYRHLASGDAVQVKLSSLNCKSKFWMNTKIAWDLPKFPPV 453  
 66 AEWVNDAYAIXYIYORPEYRHLASGDAVQVKLSSLNCKSKFWMNTKIAWDLPKFPPV 125  
 QY 454 EPPAAWGEERNVGTGLCADTKHAGLGSPLLEGCVRGREGAAANNMQVFTPTWREDIRP 513  
 126 EPPAAWGEERNVGTGLCADTKHAGLGSPLLEGCVRGREGAAANNMQVFTPTWREDIRP 185  
 QY 514 GDPQHKKCFPDAHSITSPVLYDCMSKMGQLWQKVKDKYLYHVGSGCMDCSBDHRI 573  
 186 GDPQHKKCFPDAHSITSPVLYDCMSKMGQLWQKVKDKYLYHVGSGCMDCSBDHRI 245  
 QY 574 FMNTCPSSLTQWIFEHNTNSTVLEKFRN 603  
 246 FMNTCPSSLTQWIFEHNTNSTVLEKFRN 275  
 XX

RESULT 13  
 ADD17613  
 ID ADQ17613 standard; protein; 276 AA.  
 XX  
 AC ADQ17613;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human soft tissue sarcoma-upregulated protein - SEQ ID 430.  
 XX  
 KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004048938-A2.  
 XX  
 PD 10-JUN-2004.

XX  
 RESULT 14  
 ABC08189  
 ID ABC08189 standard; protein; 335 AA.  
 XX  
 AC ABC08189;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #8180.  
 XX  
 Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX  
 food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.

PN	WO200175067-A2.
XX	11-OCT-2001.
PD	
XX	
PR	30-MAR-2001; 2001WO-US008631.
XX	31-MAR-2000; 2000US-00540217.
PR	23-AUG-2000; 2000US-00649167.
XX	(HYSE-) HYSEQ INC.
PA	DeManac RT, Liu C, Tang YT;
XX	WPI; 2001-639362/73.
DR	N-PSDB; AAS72376.
PT	New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.
XX	Claim 20; SEQ ID NO 38548; 103pp; English.
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight marker and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABC00010-ABC30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WPO at ftp.wipo.int/pub/published_pct_sequences
CC	XX
CC	Sequence 335 AA;
CC	Query Match 42.3%; Score 1385.5; DB 4; Length 335;
CC	Best Local Similarity 87.5%; Pred. No. 2.7e-132; Mismatches 253; Conservative 0; Mismatches 1; Indels 35; Gaps 1;
CC	Matches 253; Sequence 335 AA;
CC	409
CC	350 SFKYWMCGGERMEDICCSRVGHYIYRKVYKPVAGYLARNLKRVATWMDYAEYVQRR
CC	47 SHKVMWCGGRMEDICCSRVGHYIYRKVYKPVAGYLARNLKRVAVSWMDYAEYVQRR 106
CC	410 PEYRHLASGDAVQKGRSSLNCKSFKWMKTLAQLDLPKFPVPPAANGE----- 462
CC	107 PEYRHLASGDAVQKGRSSLNCKSFKWMKTLAQLDLPKFPVPPAANGEGORQSMT 166
CC	463 -----IRAVGTMGCACTRHGALGSPLRLEGVRGRGE 494
CC	167 YSKGFMFSRRFQGMVPGPPAPESPFGMGRVGTSLCADTRKGAGLGSPLRLEGVRGRGE 226
CC	495 AAWMIMQVFTTWRDIRPDPQHCKPCFDAISHTSPVLYDCHSMKGNLWKRKDT 554
CC	227 AAWMIMQVFTTWRDIRPDPQHCKPCFDAISHTSPVLYDCHSMKGNLWKRKDT 554
CC	555 LYHPVSGSCMDCSSESDHRIFANTCPSLTOQWLFHENTNSTLKENRN 603
CC	287 LYHVSGSCHDCSSESDHRIFANTCPSLTOQWLFHENTNSTLKENRN 335
CC	Sequence 269 AA;
CC	Query Match 41.8%; Score 1370; DB 4; Length 269;
CC	Best Local Similarity 100.0%; Pred. No. 7.2e-14; Mismatches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC	Matches 256; Sequence 269 AA;
CC	157
CC	98 MTPDABRVDQAYRENGFNTIVYPSDKISLNNSIPDTHPNCSKRYLSTLPNTNSIIPFFNED
CC	1 MTDABRVDQAYRENGFNTIVYPSDKISLNNSIPDTHPNCSKRYLSTLPNTNSIIPFFNED 60
CC	158 WSSLURTHSVLNRSPPPEVLTIVLVDDEHREHKKPFDYMAFPSPVILRTKGREG 217
CC	61 WSSLURTHSVLNRSPPPEVLTIVLVDDEHREHKKPFDYMAFPSPVILRTKGREG 120
CC	218 IRTMLGASVATGVVITFLDSRCEANVNPFLPLDTRARNRKTIVCPMIVDHDPRYE 277
CC	121 IRTMLGASVATGVVITFLDSRCEANVNPFLPLDTRARNRKTIVCPMIVDHDPRYE 180
CC	278 TOQADGMAFGADMEMYRIPPELQKADPSDFESPSPMAGGIYAVDPRKFWLGGYDP 337

QY 338 GLETWGGEQYEVISPKV 353  
 Db 241 GLETWGGEQYEVISPKV 256

RESULT 16

ABB66873  
 ID ABB66873 standard; protein; 666 AA.  
 XX  
 AC ABB66873;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 27411.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PP 23-MAR-2001; 2001WO-US009231.  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.

XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PT Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.

XX  
 DR N-PSDB; ABL10976.

XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.  
 PT  
 PS Disclosure; SEQ ID NO 27411; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL10976-ABL3051), expressed DNA sequences (ABL01840-ABL6175) and the encoded proteins (ABL5773-ABL2072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published\\_pct\\_sequences](ftp://wipo.int/pub/published_pct_sequences)  
 XX  
 SQ Sequence 666 AA;

Query Match 41.5%; Score 1360; DB 4; Length 666;  
 Best Local Similarity 49.0%; Fred. No. 3.3e-129; Mismatches 277; Conservative 76; Indels 52; Gaps 15;

QY 67 DGQKUKDWHIDKKEAIRDAQWVNGEQGRPPMIDAEVDOAHR--ENGENIYSDKISL 123  
 Db 121 DASVCKDWHYTFWMDKAKVGLGGKASTLSDQDIDKRSLENGNANLSDIV 180  
 QY 124 NRSLDPIRHPWNSKRYLETPNTSITIPPMENEGWSSLARTVHSVNLNSPPELVAVIV 183  
 Db 181 NRSVPDPIRKPKKEYVAKLPTVSVILVYNEYVSVLMSVHSLSINRSPPELMKILV 240  
 QY 184 DDPSPREHLKKPLEDWA-LFSPSVRLTRKREGRLRTMAGASVATGQVTFDLSCEA 242  
 Db 241 DDHSRBEVYLGKELEYTIAEHRKWRVVRPLPRTGLIGARAGARNATAEVILFDLHVEA 300  
 QY 243 NVWMLPLDIBIARAKTICPMIDVHDOPRFYETQAGDMRGAEDWMTYKRIPIPE 302

Db 301 NYWMLPLDIEPILANKRIVAVCPFDIVDHTNFHYRAQ-DEGARGAFDPEFFYKRLPLPE 359  
 QY 303 LQKADPSPDPESPWAGGLFAUDVRKWFENBLGGDGPGLTWWGQYELSPKVMCGGRMED 362  
 Db 360 DLK-HPAPDFPKSPIMAGGLPALSREFPWLGGLDINGEQYELSSFKIWWCGGENID 418

QY 363 IPCSRRVGHYR---KIVPYKYPAGVSLARNIKRVAQWMDSYAETYRER-B-YRILS 416  
 Db 419 APCSRIGHIYRGPRNHPSPRK--GDXIAHKNTRKVAEVWMDERKVNLYSHGDGLYESVD 475

QY 417 AGDVAVQKKLRSJNCKSKRKFWMTKIANDLPKYPPIPPAAMGERRNVTG-GLCAGT- 474  
 Db 476 PGDLTEQKAIKTRKUNCKAFKWFMEBEVADIMKTYPPVUPPSYANGALQJNGNQNLCLDTL 535

QY 475 ---RKGALG-----SPLRLEGCVGRGRGEAAWNMQVFTFWREDIRPGDPQHTCKF 522

Db 536 GRKKIINKGKMGYACADNIKTPQR-----TQFWELSWKRLR---LRRGE 576

QY 523 CFDA--ISITSPVTLYDPEHSMKSNOLWRY--RKOKTLYPVSQG-SCMPSBESDRIFWMTC 578  
 Db 577 CLDVQIWADANAPWMLWDCHSOGGNQWYQWYDVRQKOLKGTEGRRCLELLPFSQEVANKC 636

QY 579 NPSSITQOMLPEKINSTUERKFNR 633  
 Db 637 DTDRFQDQINFGSPNKTALDNYSD 651

RESULT 17

ABB58867  
 ID ABB58867 standard; protein; 666 AA.  
 XX  
 AC ABB58867;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 3393.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 PR 27-SEP-2001.  
 PR 23-MAR-2001; 2001WO-US009231.  
 PR 11-JUL-2000; 2000US-00614150.

XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PT Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.

XX  
 DR N-PSDB; ABL02970.

XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.  
 PT  
 PS Disclosure; SEQ ID NO 3393; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL10976-ABL3051), expressed DNA sequences (ABL01840-ABL6175) and the encoded proteins (ABL5773-ABL2072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly

CC from WIPO at [ftp://wipo.int/pub/published\\_pct\\_sequences](ftp://wipo.int/pub/published_pct_sequences)  
 XX SQ Sequence 666 AA;  
 Query Match Best Local Similarity 41.5%; Score 1360; DB 4; Length 666;  
 XX Matches 277; Conservative 76; Mismatches 160; Indels 52; Gaps 15;  
 QY 67 DQGKUKDMDKFAIRRDAQVNGEQGRYRPMIDAERVDQAYR---ENGFTIVYSDKSL 123  
 DB 121 BASVYKDKHDYTFMEKDAKRVGIGLEGKASTLDESDQRBLKRMVSLENGFLNALSISV 180  
 QY 124 NRSALPDIRHNCNSKRYLETLPTNTSIIIPHNEGSSLRTVHSVLNSPPPELVAEVV 183  
 181 NRSVPDIRPLCRKKEVAKPLPTSVLIVIFNEYLVSILMSVSLNSPPPELMEKILV 240  
 QY 184 DDFSDREHKKPLKEDYMA-LFSPVRLRURKKEGLRTRMGLASGVATGVTIFDSHEA 242  
 DB 241 DHDSDREYLGKELETYIAHFKWVVRVLRPTGTLIGARAGNARATAVLFLSHVEA 300  
 QY 243 NVMWLPPLDILARAKRTIVCPMDVHDPPDPRYETQDAMGAPDMYKRPIDP 302  
 301 NYNNLPLPLPEPIALANKRTAVCPFDVIDHTNFHYRAQ-DEGARGAFDPEPFYKRULP 359  
 QY 303 LOKADPSPDPSPSPWMAAGIFPAVDRKWFALGIDPGLTETWGEODEYELSKFVWMMGGRNED 362  
 DB 360 DLK-HPAPFSPIMAGGUFAISREFFWFLGEGDGLWGEQYELSPKIMWGGEMD 418  
 QY 363 IPESRVGHYR----KVPYKVAGVSLARNLKRAETVMDAYEAYIYORRSE-YRILS 416  
 DB 419 APECSRIGHTYGRPRHNPSPRK--GDYLIHKNYKGRVABEYWMDSKKNYLYSHGDLYESD 475  
 QY 417 AGYVAVOKKURSSUNCKSKWFMKIAALDPLKPVPPVERPAAMGEBIRNGT-GLCAET- 474  
 DB 476 PGDLTEOKAIRTKLNUCKSKFWMFEEVAFDLMKTYPPVDPERSYANGALONGNOMICLDTL 535  
 QY 475 --KHALG-----SPLRLGCVRGVRGEAANNMVOFTFWREDITRGDPOHTK 522  
 DB 536 GRKKGHNKGWMYACADNIKUPQR-----TQFWLMSWKDRU---LRRKG 576  
 QY 523 CFDA--ISHTSPVTLYDCHSMKGNGQNLWQY-KRDKTLYHPSG-SOMCDSESHDRIFMTC 578  
 DB 577 CLDVOIWDANAAPWLVWDCHSQGGNQYWTYDRHKQKHGTEGRKRCLELLPPSOEVANKC 636  
 QY 579 NPSSLTQMLFEHTNSTVLEKFNEN 603  
 DB 637 DTDNRFQOQNFGSFNKTAIDNYSOD 661

RESULT 18  
 AAU0774  
 ID AAU0774 Standard; protein; 321 AA.  
 AC AAU0774;  
 XX DT 04-DEC-2001 (first entry)  
 XX DE Human novel transferase protein, NHP #17.  
 XX KW Human; transferase; breast cancer; prostate cancer; immunogen; gene therapy; antisense.  
 OS Homo sapiens.  
 PN WO200164903-A2.  
 XX PD 07-SEP-2001.  
 XX PR 28-FEB-2001; 2001WO-US006460.  
 XX PR 29-FEB-2000; 2000US-0185920P.  
 PR 02-MAR-2000; 2000US-0186558P.  
 PR 24-MAR-2000; 2000US-0191849P.

XX (LEXI-) LEXICON GENETICS INC.  
 PA XX  
 PI Donoho G, Hilbun B, Turner CA, Friedrich G, Abuin A; Hu Y, Kieke JA;  
 PI Zambrowicz B, Sands AR, Walke DW, Wilgnowski NL, Potter DG;  
 PI DR WPI; 2001-550185/61.  
 DR N-PSDB; AAS12621.  
 XX PT Novel nucleic acid sequences encoding novel human proteins useful for  
 PT diagnosis, drug screening, clinical trial monitoring and treatment of  
 PT diseases and disorders.  
 XX Disclosure; Page 50-51; 60pp; English.  
 CC The invention relates to isolated nucleic acids and proteins are useful for  
 CC transferase proteins (NHP). The nucleic acids and proteins are useful for  
 CC diagnosis, drug screening, clinical trial monitoring and treatment of  
 CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be  
 CC useful for augmenting the efficacy of chemotherapeutic agents used in  
 CC treatment of breast or prostate cancer. The nucleic acid is also useful  
 CC in NHP gene regulation, and as antisense primers in amplification  
 CC reactions of NHP gene sequences. NHPs are useful for producing  
 CC antibodies. The present sequence represents a novel human transferase  
 XX which has sequence similarity to N-acetyl-galactosaminyltransferase  
 SQ Sequence 321 AA;

Query Match Best Local Similarity 40.0%; Score 1311; DB 4; Length 321;  
 XX Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MRRIKEKRLQIAVAVLAVLAVLVLNGLWYRERQDPCTPGCSGAAVPAAGCSSHSDQ 60  
 DB 1 MRRIKEKRLQIAVAVLAVLAVLVLNGLWYRERQDPCTPGCSGAAVPAAGCSSHSDQ 60  
 QY 61 KTFFLGDCOKLQKWHDKAIRRDAQVNGEORPYRPMIDAERVDQATRENGNIYVSDK 120  
 DB 61 KTFFLGDCOKLQKWHDKAIRRDAQVNGEORPYRPMIDAERVDQATRENGNIYVSDK 120  
 QY 121 ISLNSLSDIRHNCNSKRYLETLPTNTSIIIPHNEGSSLRTVHSVLNSPPPELVI 180  
 DB 121 ISLNSLSDIRHNCNSKRYLETLPTNTSIIIPHNEGSSLRTVHSVLNSPPPELVI 180  
 QY 181 VLVDDFSRBHLKPLDYMALFSPVRLRURKKEGLRTRMGLASGVATGVTIFDSHC 240  
 DB 181 VLVDDFSRBHLKPLDYMALFSPVRLRURKKEGLRTRMGLASGVATGVTIFDSHC 240  
 QY 241 BANVNWLPPLDR : 253  
 DB 241 BANVNWLPPLDR : 253

RESULT 19  
 ABB59363  
 XX ABB59363 Standard; protein; 1252 AA.  
 XX AC ABB59363;  
 XX DT 26-MAR-2002 (first entry)  
 XX DB Drosophila melanogaster polypeptide SEQ ID NO 4881.  
 XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX OS Drosophila melanogaster.  
 XX PN WO200111042-A2.  
 XX PD 27-SEP-2001.



PT	rheumatoid arthritis, lupus, fibrosis, osteoarthritis, fibromyalgia and gout.
PS	Disclosure; Page: 56pp; English.
XX	The invention relates to a novel method for diagnosing and analysing patient sample containing mRNA, analysing gene expression using the mRNA that results in a gene expression signature of the mRNA, and using that gene expression signature to diagnose or analyse the autoimmune disease or arthritides in the patient, where gene expression of at least 60% of the genes correlates with that of the gene signature. The invention further comprises: a treatment of rheumatoid arthritis; identification of genes for targeting in the treatment of rheumatoid arthritis in a mammal other than a mouse; diagnosis of rheumatoid arthritis in a mammal; an array or gene chip, specific for rheumatoid arthritis; diagnosis or analyses of autoimmune disease or rheumatoid arthritis; screening the efficacy of a candidate drug in vitro for the treatment of collagen-induced arthritis; and reducing the symptoms associated with collagen-induced arthritis. The compositions of the invention have the following activities: immunosuppressive, antirheumatic, antiarthritic, osteopathic, antigenic, antiinflammatory, dermatological, and immunomodulatory. The methods and compositions of the present invention are useful for diagnosing and treating autoimmune disease or arthritides, such as rheumatoid arthritis, lupus, ankylosing spondylitis, fibrosis, fibromyalgia, osteoarthritis, gout, juvenile rheumatoid arthritis, and an immune disease caused by an infectious agent. This sequence represents a protein sequence relating to the genes used in the analysis and treatment of autoimmune diseases or arthritides. Note: This sequence is not shown in the specification. It has been supplied in an electronic format from WIPO.
CC	RESULT 21
CC	Query Match 34.3%; Score 1125; DB 7; Length 559;
CC	Best Local Similarity 44.0%; Pred. No. 2.8e-105;
CC	Matches 232; Conservative 87; Mismatches 170; Indels 38; Gaps 16;
CC	Query 88 GNGEQGRP-YPMDAERUDQAYRENGNIYVSDKISLRSRSLPDRPHPNQNSKRYLETP 145
CC	DB 57 GPGBMGKPKVPTPKEDOKKEMKFNOFLNMASEMIALNRSPLPDRPLGCKTKVYPDNL 116
CC	Query 146 NTSTIIPFHNGNSLLRTVHSVNRSPPELVAELIVLVDPSDRHLKPLKEDYM-ALFP 204
CC	DB 117 TTSVIVVPHNEANSTLRTVHSVNRSPHMISLVLVDDASERDFLKPPLPESYKGLK 176
CC	Query 205 SVRLRTKREGIIRTRMIGASVATGVDVITPLSHCEANNWLPPLDRJARNRKTIVC 264
CC	DB 177 PVAVIRMQRGLRARIKAGAVSKQVITFLDHCETCWTGWLBPDLARKRHDRTVCP 236
CC	Query 255 MIDVHDPRYETQAG-DAMRGADPMWYKRIPIP--ELOKADPSPDPESPVMAG 320
CC	DB 237 IIVDSDTTEY--MAGSDMTYSGFNWKLNFRWYVPUOREMDRKGDRTLFVPTPMAGG 294
CC	Query 321 LFAVDRKPHWBLGGYDPGLIWIWGGCQEYTSFKVWPGGMDIPCSRVSQVGHLYRKVYPKV 380
CC	DB 295 LFSIDRQFOEIGTYDAGMDIWGGENLEISPRWQCGGTEIVCWSHVFRKATPYF 354
CC	Query 381 PAGVS-LARRNKRVAEVWMDYAEVYVORRPEYRHLSSADVAVOKLKSLSNCKSPKWF 438
CC	DB 355 PGGGQQLINNNRKLAEVWDEFFNNFFYIISPGVTKVDYGDISERVGRLKQCKPFWY 414
CC	Query 439 MTKLAWD-LPKYVPPVPPAAMGEIRNGTGLCAPDTKGALGSPRLLEGCVVRGEAA 496
CC	DB 415 LENYIPDSDQIPRY-----FSLGEIRVNETQCLDNMRKENEKVGFFNC-HGMG--- 463
CC	Query 497 WNNMQVFTFTWREDIREPGDQPHQTKFCDAISHTSPVTLIDCHMKGNOLWVKYKDK-TI 555
CC	DB 464 -GNQVFSYITANKEIRTD----LCIDVSKLNGPVTMLKCHLKGNOUWEPVPLK 515
PT	PT Diagnosing and analyzing autoimmune disease using gene expression profiles and microarray technology, useful for diagnosing and treating
PT	PT WPI: 2003-712740/67.
DR	DR GENBANK: NP_065207.

RESULT 22  
 ID AAB43561  
 AAB43561 standard; protein; 561 AA.  
 AC AAB43561;  
 XX DT 08-FEB-2001 (first entry)  
 XX DE Human cancer associated protein sequence SEQ ID NO:1006.  
 XX KW Human; cancer associated gene; cancer antigen; detection; cancer;  
 KW diagnosis; cytostatic; proliferative; vulerary; immunomodulator;  
 KW antidiabetic; antilasthmatic; antiarheumatic; antiarthritic; antiviral;  
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;  
 KW dermatologic; neuroprotective; thrombolytic; nootropic;  
 KW vasoactive; antiparotic; antiangiogenic; gene therapy; inflammation;  
 KW immune disorder; haemopoietic cell disorder; autoimmune disorder;  
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
 KW neurological disease; drug screening.  
 XX OS Homo sapiens.  
 XX WO200055350-A1.  
 XX PN 21-SEP-2000.  
 XX PP 08-MAR-2000; 20000NO-US005882.  
 PR 12-MAR-1999; 99US-0124270P.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Rosen CA, Ruben SM;  
 XX DR WPI; 2000-587533/55.  
 DR N-PSDB; AAC77770.  
 XX PT Novel isolated nucleic acids comprising sequences encoding peptides  
 PT useful for treating or diagnosing e.g. cancer.  
 XX PS Claim 11; Page 1581-1584; 2352PP; English.  
 XX CC AAC77607 to AAC7848 encode the human cancer associated proteins given in  
 CC AAC77607 to AAC7848. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities  
 CC include: cytostatic; proliferative; vulnerary; immunomodulator;  
 CC antidiabetic; antiasthmatic; antiarheumatic; antiarthritic;  
 CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;  
 CC dermatologic; neuroprotective; cardiant; thrombolytic; coagulant;  
 CC nootropic; vasoactive; antiparotic; antiangiogenic. The  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polypeptides, polypeptides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells to treat disorders of haemopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC refection, modulate haemostatic or thrombolytic activity, modulate  
 CC inflammation, cancers, cardiovascular disorders, neurological disease and  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AAC7849 to  
 CC AAC78457 and AAC78440 represent sequences used in the exemplification of  
 XX the present invention.  
 SQ Sequence 561 AA;

Query Match 34.3%; Score 1125; DB 3; Length 561;  
 Best local Similarity 44.0%; Pred. No. 2-88-105;  
 Matches 222; Conservative 87; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGBQGR--YPMDAEVDQAYRERENPVIYSDKISLNRSIDPDRHNCNSKRYLETLP 145  
 59 GPGEMGKPVVIRPKEDQEKMFINGKONLMASEMIAANRSLPDVRLGCKTKVPOMLP 118  
 DB 146 NTSTIIPPHNEGSSLTAVHSLRSPPELAVIIVUDFSDREKJKPIEDM-AIIFP 204  
 QY 119 TTSVIVVHNEAMSTLLTVHSVJNRSRHMTEIIVUDASERDFLRPLESYVKKV 178  
 DB 205 SVRTIRTKEGRJTRMAGASVATGDTYTFDLSHCEANNNWPLPLRIARNRKTIVCP 264  
 DB 179 PVRVIRMEORGSLIRAKLGAVSKQVITFLDANHCTVGMLPULLARIKHDRRTVCP 238  
 QY 265 MIDVHDHPFVYTOAG-DAMRAFDWNYKRPPIP---ELOKADPSDPSPVMMG 320  
 DB 239 IIDVSDDFEY--MAGSDMTYGFNPWNLNFVWVPPREMDRKGRDPVLPKPTMGG 296  
 QY 321 LFAVDRKFWELGGYDPPLAEWKGQVVISPKWMCGRMEDIPCSRVGHTYKVYKPV 380  
 DB 297 LFSIDRDFYFOEIGYDAGMDINGENLHISFRIWOCGCTLEIVTCSHVGHVPRKATPYTP 356  
 QY 381 PAGVS-LARNLKEVAEWNDEAYIYQRRPBYRHISAGDVAVOKKRSIINCKSFKWP 438  
 DB 357 PGGCQIQINNNRLLAEWMDERKQFPLKISPGVTKYGDSSRVGARHKQCKPSSWY 416  
 QY 439 MTKIAWD-LPKFVPPVPPAAMWGEINNGTICADPKHGAALGSPKLULEGCVRGRAA 496  
 DB 417 LENIYPDSQIPRHY-----FSIGBIRNVENTQCDLMARKENKEQIIFNC-HGMG--- 465  
 QY 497 WANNQVFRTWREDPDRDQHKKFCDAISHTSPVLYDCHSMKGQNLWQYRKDK-TL 555  
 DB 466 --GNQVFSTYANKBIRTD-----LCUDVSKUNGPNMLKCHHLKGQNLWEDPVKUL 517  
 QY 556 YHPVSGSCMD-CSFSDHRI-FMTCNPSSLTQMLFENTNSTLKEP 600  
 DB 518 QHVNQNQCIDKATEEDQVSPSRDCN-GSRSQWLLR--NVTLEIF 561

RESULT 23  
 ID ADB79836  
 ID ADB79836 standard; protein; 559 AA.  
 AC ADB79836;  
 XX DT 04-DEC-2003 (first entry)  
 XX DE Rat polypeptide GalNAc transferase T1, SEQ ID 76.  
 XX KW Analgesic; pain; streptozocin-induced diabetes; rat.  
 XX OS Rattus norvegicus.  
 XX EP1279744-A2.  
 XX PD 29-JAN-2003.  
 XX 26-JUL-2002; 2002EP-00255249.  
 XX 27-JUL-2001; 2001GB-00018354.  
 PR 07-FEB-2002; 2002GB-00002910.  
 XX (WARN ) WARNER LAMBERT CO.  
 XX PPI Brookbank RA, Dixon AK, Lee K, Pinnock RD;  
 XX WPI; 2003-395407/38.  
 DR N-PSDB; ADB79837.  
 XX  
 SQ Use of isolated gene sequences and encoded polypeptides that are  
 PT upregulated in the spinal cord in response to streptozocin-induced  
 PT diabetes for screening compounds for the treatment of pain, or for  
 PT diagnosing pain.  
 XX  
 Claim 1; Page 148-149; 334PP; English.

CC the present invention relates to nucleotide sequences which are useful in  
CC the screening of compounds for the treatment of pain, or for the  
CC diagnosis of pain. The nucleotide sequences are up-regulated in the  
CC spinal cord in response to streptozocin-induced diabetes. The present  
CC sequence was used to illustrate the invention.  
XX

卷之三

Query Match      34.0%; Score 1116; DB 7; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 2; Jc=104;  
 Matched 230. Comparisons 99. Mismatches 171. Inverts 39. Gaps 162.

Brooksbank RA, Dixon AK, Lee K, Pinnoch RD; WPI; 2003-395407/38. N-FSDB; ADB79797.

Use of isolated gene sequences and encoded polypeptides that are upregulated in the spinal cord in response to streptozocin-induced diabetes for screening compounds for the treatment of pain, or for diagnosing pain.

Claim 1; page 91-93; 33app; English.

The present invention relates to nucleotide sequences which are useful in the screening of compounds for the treatment of pain, or for the diagnosis of pain. The nucleotide sequences are up-regulated in the spinal cord in response to streptozocin-induced diabetes. The present sequence was used to illustrate the invention.

卷之三

Query Match Score: 34.07; Length: 559;  
 Best Local Similarity Score: 11.23; DE-104;  
 Matches: 230; Conservative: 43.6%;  
 Mismatches: 88; Deletions: 171; Insertions: 38; Gaps: 16;

Qy	Db
88 GNGEGRG--YPTMDAERVDOAERENGNIYVSDKISKLRSRLPPIRHPNCNSKRLTLETPL 145	57 GGEGMGEKVVVIKEDQERKMKWKLINQFILMASEMIAFNSLDPVLEGKTKVYPPDSL 116

Db 237 IIVIVSDDTFBY--MASSDMTQGPNWMLNFRMVPVOREMDRKGDTLPVRPTMMAGG 294  
 Qy 321 LIJAVDRKQFMWELGGYDPLBINGGEQYISFKYTMGGERMDIPCSRVRHIIYKRVYKVY 380

Db	QY
295	LISIDRRYFOEIGTYAGMDINGGENLBISFRITWOGGLEIIVCSHGVGFVRKATPYT
381	PAGVS - LARNJLKRAEVWMDBEAFLYQRRPEYRHLASGDAVAVQKKKLASSLNCEFSKWP 438

Db  
355 PEGTQQLINKNHRLLAEVMDPESKNNFFYITISPGVTKVYGDISRVGLRHKLQCKPFSWY 414  
Ov 439 MTKIAAD--LRPKYPPTEPPAAMGEIIRNGTGLCQDTRKGAGSPRLRGCVGRGERGAA 496

Db	Qy	556	559
		YIPVUSCMB--CSESDIRI--FKMTCNPSLLOI  WLFEMINSTVAKP OHVNSNGCLDKATEBDSQVSPRDRCT--GSRSQWLRL--NVTUPEIP	600

RESULT 25  
ADE65526

IB ADB62526 Standard; protein; 559 AA.  
 XX  
 AC ADB62526;  
 XX

XX  
PR 27-JUL-2001; 2001GB-00018354.  
PR 07-FEB-2002; 2002GB-00002910.  
XX

Search completed: November 22, 2004, 13:41:16  
Job time : 74 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

Om protein - protein search, using sw model  
Run on: November 22, 2004, 13:39:57 ; Search time 21 Seconds

(without alignments)  
1904.274 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278

Sequence: 1. MRRKKRLLQVALVIAALV.....TQQLFERTNSTVLEKFFNRN 603

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 20000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/iaa/5a\_COMB.pep:\*

2: /cgn2\_6/ptodata/1/iaa/5b\_COMB.pep:\*

3: /cgn2\_6/ptodata/1/iaa/6a\_COMB.pep:\*

4: /cgn2\_6/ptodata/1/iaa/6b\_COMB.pep:\*

5: /cgn2\_6/ptodata/1/iaa/PCRTS\_COMB.pep:\*

6: /cgn2\_6/ptodata/1/iaa/backfile1.pep:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	3278	100.0	603	4	US-03-795-926-43	Sequence 43, Appl
2	3242.5	98.9	631	4	US-03-795-926-41	Sequence 41, Appl
3	2771	84.5	506	4	US-03-795-926-31	Sequence 31, Appl
4	2746.5	83.8	535	4	US-03-795-926-29	Sequence 29, Appl
5	1877	57.3	365	4	US-03-795-926-39	Sequence 39, Appl
6	1370	41.8	269	4	US-03-795-926-27	Sequence 27, Appl
7	1311	40.0	321	4	US-03-795-926-35	Sequence 35, Appl
8	115	34.0	517	3	US-03-967-508-19	Sequence 19, Appl
9	1115	34.0	517	5	PCT-US94-02552-19	Sequence 19, Appl
10	1115	34.0	517	2	US-03-795-926-19	Sequence 19, Appl
11	1115	34.0	559	2	US-03-967-508-9	Sequence 9, Appl
12	1115	34.0	559	3	US-03-967-506-9	Sequence 9, Appl
13	1115	34.0	559	5	PCT-US94-02552-9	Sequence 9, Appl
14	1026	31.3	578	4	US-03-217-306B-2	Sequence 2, Appl
15	1024	31.2	560	4	US-03-217-306B-10	Sequence 10, Appl
16	990	30.2	240	4	US-03-795-926-33	Sequence 33, Appl
17	988	30.1	209	4	US-03-795-926-37	Sequence 37, Appl
18	971	29.6	552	4	US-10-140-002-196	Sequence 196, Appl
19	941	28.7	302	4	US-03-270-767-33733	Sequence 33733, Appl
20	926	28.2	639	3	US-03-376-852-2	Sequence 2, Appl
21	920	28.1	638	3	US-03-347-488-2	Sequence 10, Appl
22	916	27.9	633	2	US-03-648-298-2	Sequence 33, Appl
23	804	24.5	224	4	US-03-795-926-23	Sequence 37, Appl
24	744	22.7	603	4	US-03-973-457-2	Sequence 2, Appl
25	675	20.6	289	4	US-03-270-767-45334	Sequence 45334, Appl
26	561	17.1	238	4	US-03-270-767-46411	Sequence 46411, Appl
27	503.5	15.4	4	US-03-270-767-43543	Sequence 43543, Appl	

28 483 14.7 143 4 US-09-795-926-21  
29 481 14.7 112 4 US-09-795-926-25  
30 450.5 13.7 149 4 US-09-795-457-6  
31 428.5 13.1 161 4 US-09-270-767-58910  
32 287 8.8 122 4 US-09-513-995C-7570  
33 277.5 8.5 148 4 US-09-270-767-3287  
34 241.5 7.4 313 4 US-09-270-767-46209  
35 234.5 7.2 300 4 US-09-270-767-43809  
36 228.5 7.0 290 4 US-09-270-767-61793  
37 218.5 6.7 249 4 US-09-270-767-33350  
38 168 5.1 73 4 US-09-270-767-48818  
39 168 5.1 73 4 US-09-270-767-48817  
40 167 5.1 101 4 US-09-270-767-33441  
41 167 5.1 101 4 US-09-270-767-48653  
42 133 4.1 424 4 US-09-131-000-5947  
43 129 3.9 88 4 US-09-270-767-33601  
44 129 3.9 88 4 US-09-270-767-48818  
45 128 3.9 135 4 US-09-131-000-5086  
46 127 3.9 436 4 US-09-328-352-4416  
47 125.5 3.8 955 4 US-09-437-277-3  
48 124 3.8 702 4 US-09-134-000-5087  
49 123 3.8 534 4 US-09-134-000-5087  
50 123 3.8 721 4 US-09-107-5320-6889  
51 121.5 3.7 1055 4 US-09-134-000-5086  
52 119.5 3.6 492 2 US-08-460-812-7  
53 119.5 3.6 492 2 US-08-238-832-7  
54 119.5 3.6 492 3 US-08-590-563-7  
55 119.5 3.6 492 4 US-09-770-621-4  
56 119.5 3.6 492 4 US-09-770-621-7  
57 119.5 3.6 492 4 US-08-770-621-7  
58 119.5 3.6 492 4 US-08-460-812-5  
59 119.5 3.6 492 4 US-09-238-832-7  
60 118 3.6 480 3 US-08-590-563-5  
61 118 3.6 480 4 US-09-770-621-5  
62 118 3.6 480 4 US-09-770-621-8  
63 118 3.6 480 4 US-09-235-832-5  
64 116.5 3.6 88 4 US-09-270-767-33890  
65 116.5 3.6 88 4 US-09-270-767-49107  
66 116.5 3.6 909 4 US-09-425-383-2  
67 111.5 3.4 491 2 US-08-460-812-8  
68 111.5 3.4 491 3 US-08-590-563-8  
69 111.5 3.4 491 4 US-09-770-621-8  
70 111.5 3.4 491 4 US-09-235-832-8  
71 107.5 3.3 507 4 US-09-130-3374-25  
72 107 3.3 270 4 US-09-810-0284-39  
73 107 3.3 270 4 US-10-303-162-39  
74 107 3.3 270 4 US-09-495-406-25  
75 105.5 3.2 187 4 US-09-495-406-25  
76 105.5 3.2 187 4 US-09-475-457-4  
77 105 3.2 336 4 US-09-542-681A-6535  
78 105 3.2 127 1 US-08-392-828C-39  
79 105 3.2 127 1 US-09-330-945-39  
80 105 3.2 324 1 US-08-591-236-10  
81 105 3.2 324 1 US-08-741-682A-10  
82 102.5 3.1 540 1 US-08-485-286-77  
83 102 3.1 317 4 US-09-583-110-4849  
84 101.5 3.1 395 4 US-08-635-552A-4  
85 101.5 3.1 419 2 US-08-270-581-2  
86 101.5 3.1 419 4 US-09-140-859-3  
87 101.5 3.1 540 1 US-08-675-499A-5  
88 101.5 3.1 419 4 US-08-812-008-5  
89 101 3.1 328 4 US-09-583-110-3554  
90 100 3.1 331 4 US-09-634-224-247  
91 99.5 3.0 303 4 US-09-495-406-17  
92 99.5 3.0 303 4 US-09-810-0284-29  
93 99.5 3.0 303 4 US-10-303-162-29  
94 99 3.0 341 4 US-09-328-352-7178  
95 98 3.0 301 4 US-09-495-406-15  
96 98 3.0 301 4 US-09-816-0284-27  
97 98 3.0 301 4 US-10-303-162-27  
98 96.5 2.9 623 1 US-08-653-740-7  
99 96.5 2.9 623 2 US-09-273-594-7  
100 96.5 2.9 623 2 US-09-275-925-7

## ALIGNMENTS

RESULT 1

US-09-795-926-43 ; Sequence 43, Application US/09795926

PATENT NO. 6555669 ; GENERAL INFORMATION:

APPLICANT: Donoho, Gregory

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Friedrich, Glenn

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sands, Arthur T.

APPLICANT: Walk, D. Wade

APPLICANT: Hu, Yi

APPLICANT: Kieke, James Alvin

APPLICANT: Porter, David George

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND POLYNUCLEOTIDES ENCODING THE SAME

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795, 926

CURRENT FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: US 60/185, 920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186, 558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191, 849

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: PastSeq for Windows Version 4.0

SEQ ID NO: 43

LENGTH: 603

TYPE: PRT

ORGANISM: homo sapiens

US-09-795-926-43

Query Match 100.0%; Score 3278; DB 4; Length 603;

Best local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 603; Conservative 0;

QY 1 MRRKEKRLIQQAVLAVALVLAALVLPNGVGLWLYRERQDPGPGGSGAAVAPAGOSSHRSQK 60

Db 1 MRRKEKRLIQQAVLAVALVLAALVLPNGVGLWLYRERQDPGPGGSGAAVAPAGOSSHRSQK 60

QY 1 KTFPLGDGOKLKDWDKEAIRRDAQGVNGEQRGPYPMDAERVDQAYRENGNIYVSDK 120

Db 61 KTFPLGDGOKLKDWDKEAIRRDAQGVNGEQRGPYPMDAERVDQAYRENGNIYVSDK 120

QY 121 ISLNSLSPDTIRPHNCGSKYKLETFNTSIIIPFNEGWSLLRTVHSVLRSPPELVAEI 180

Db 121 ISLNSLSPDTIRPHNCGSKYKLETFNTSIIIPFNEGWSLLRTVHSVLRSPPELVAEI 180

QY 181 VLVDDPSDREHLKPLEDYLALFSPVSRVIRAKTKREGLIRTRMIGASVATGVDVTPLDSC 240

Db 181 VLVDDPSDREHLKPLEDYLALFSPVSRVIRAKTKREGLIRTRMIGASVATGVDVTPLDSC 240

QY 241 EANVWNLPLLDIARNRKTCVCPMDVDDHDPRTQAGDAMGAFDMEVYKIRIP 300

Db 241 EANVWNLPLLDIARNRKTCVCPMDVDDHDPRTQAGDAMGAFDMEVYKIRIP 300

QY 301 FELQKADPSDFPESVPMAGGLFAVDRKWFELLGDPGLIWGSGQYEVLSFKVWMCGRM 360

Db 301 FELQKADPSDFPESVPMAGGLFAVDRKWFELLGDPGLIWGSGQYEVLSFKVWMCGRM 360

QY 361 EDIPSRVGHYKVKYVKYVPGVSLARNKRCVAVWMDYAYKQRPYRILSAGDV 420

Db 361 EDIPSRVGHYKVKYVKYVPGVSLARNKRCVAVWMDYAYKQRPYRILSAGDV 420

QY 481 SPLRLEGCTGRGRGEAAWNINMQVFTFWBDRIDFGDPQMKKFDALSHTSPTVLYDHS 540

Db 481 SPLRLEGCTGRGRGEAAWNINMQVFTFWBDRIDFGDPQMKKFDALSHTSPTVLYDHS 540

QY 541 MKGNOLWKYTRKDLYHPPVSGSCMDCSHSDRIFTMNTCPSSHTQWMTFENTSTVLPK 600

Db 541 MKGNOLWKYTRKDLYHPPVSGSCMDCSHSDRIFTMNTCPSSHTQWMTFENTSTVLPK 600

QY 601 NRN 603

Db 601 NRN 603

RESULT 2

US-09-795-926-41 ; Sequence 41, Application US/09795926

PATENT NO. 6555669 ; GENERAL INFORMATION:

APPLICANT: Donoho, Gregory

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Friedrich, Glenn

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sands, Arthur T.

APPLICANT: Walk, D. Wade

APPLICANT: Hu, Yi

APPLICANT: Kieke, James Alvin

APPLICANT: Porter, David George

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND POLYNUCLEOTIDES ENCODING THE SAME

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795, 926

CURRENT FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: US 60/185, 920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186, 558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191, 849

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: PastSeq for Windows Version 4.0

SEQ ID NO: 41

LENGTH: 631

TYPE: PRT

ORGANISM: homo Bapiens

US-09-795-926-41

Query Match 99.9%; Score 3242.5; DB 4; Length 631;

Best local Similarity 95.4%; Pred. No. 0; Mismatches 0; Indels 29; Gaps 1;

Matches 601; Conservative 0;

QY 1 MRRKEKRLIQQAVLAVALVLAALVLPNGVGLWLYRERQDPGPGGSGAAVAPAGOSSHRSQK 60

Db 1 MRRKEKRLIQQAVLAVALVLAALVLPNGVGLWLYRERQDPGPGGSGAAVAPAGOSSHRSQK 60

QY 61 KTFPLGDGOKLKDWDKEAIRRDAQGVNGEQRGPYPMDAERVDQAYRENGNIYVSDK 120

Db 61 KTFPLGDGOKLKDWDKEAIRRDAQGVNGEQRGPYPMDAERVDQAYRENGNIYVSDK 120

QY 121 ISLNSLSPDTIRPHNCGSKYKLETFNTSIIIPFNEGWSLLRTVHSVLRSPPELVAEI 180

Db 121 ISLNSLSPDTIRPHNCGSKYKLETFNTSIIIPFNEGWSLLRTVHSVLRSPPELVAEI 180

QY 181 VLVDDPSDREHLKPLEDYLALFSPVSRVIRAKTKREGLIRTRMIGASVATGVDVTPLDSC 240

Db 181 VLVDDPSDREHLKPLEDYLALFSPVSRVIRAKTKREGLIRTRMIGASVATGVDVTPLDSC 240

QY 241 EANVWNLPLLDIARNRKTCVCPMDVDDHDPRTQAGDAMGAFDMEVYKIRIP 300

Db 241 EANVWNLPLLDIARNRKTCVCPMDVDDHDPRTQAGDAMGAFDMEVYKIRIP 300

QY 301 FELQKADPSDFPESVPMAGGLFAVDRKWFELLGDPGLIWGSGQYEVLSFKVWMCGRM 360

Db 301 FELQKADPSDFPESVPMAGGLFAVDRKWFELLGDPGLIWGSGQYEVLSFKVWMCGRM 360

QY 361 EDIPSRVGHYKVKYVKYVPGVSLARNKRCVAVWMDYAYKQRPYRILSAGDV 420

Db 361 EDIPSRVGHYKVKYVKYVPGVSLARNKRCVAVWMDYAYKQRPYRILSAGDV 420

QY 421 AVQKLRSSINCYSPKWMKTFKIAWDLPTKPYPPRPAWGERBIRNGCGLACDTKHGIG 480

Db 421 AVQKLRSSINCYSPKWMKTFKIAWDLPTKPYPPRPAWGERBIRNGCGLACDTKHGIG 480

RESULT 3  
US-09-795-926-31

; Sequence 31, Application US/09795926

; Patent No. 655669

; GENERAL INFORMATION:

; APPLICANT: Donoho, Gregory

; APPLICANT: Hilburn, Erin

; APPLICANT: Turner, C. Alexander Jr.

; APPLICANT: Friedrich, Glenn

; APPLICANT: Abuin, Alejandro

; APPLICANT: Zambrowicz, Brian

; APPLICANT: Sands, Arthur T.

; APPLICANT: Walk, D. Wade

; APPLICANT: Wilgowski, Nathaniel L.

; APPLICANT: Hu, Yi

; APPLICANT: Kleke, James Alvin

; APPLICANT: Porter, David George

; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND POLYNUCLEOTIDES ENCODING THE SAME

; FILE REFERENCE: LEX-0144-USA

; CURRENT APPLICATION NUMBER: US/09/795,926

; CURRENT FILING DATE: 2001-02-29

; PRIOR APPLICATION NUMBER: US 60/185,920

; PRIOR FILING DATE: 2000-02-29

; PRIOR APPLICATION NUMBER: US 60/186,558

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 60/191,849

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO: 31

; LENGTH: 506

; TYPE: PRT

; ORGANISM: homo sapiens

; US-09-795-926-31

Query Match 84.5%; Score 2771; DB 4; Length 506;

Best Local Similarity 100.0%; Pred. No. 2.; ne-289; Mismatches 0; Indels 0; Gaps 0;

Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDABRVDQAYRENGFNIYUSDKISLNRSIDPDIRPHNCNSKRYELEPTNSIIIPFHNBG 157

Db 1 MTDABRVDQAYRENGFNIYUSDKISLNRSIDPDIRPHNCNSKRYELEPTNSIIIPFHNBG 60

QY 241 EANVNWLPPLDRTARNRKTIVCPMIDVIDDHDFFRYETOAGDAMRGAFDWMETKRP 300

Db 241 EANVNWLPPLDRTARNRKTIVCPMIDVIDDHDFFRYETOAGDAMRGAFDWMETKRP 300

QY 301 PELQKADSDPFPESPVMAGGLFAVDRKWNWELGSDYDGLFGEYELSK----- 352

Db 301 PELQKADSDPFPESPVMAGGLFAVDRKWNWELGSDYDGLFGEYELSK----- 352

QY 353 -----VMCGGRMEDIPCSRVGHTYRKVYKUPGVSILARN 391

Db 361 VNSNWPQAVFLPRAFPNMLAQVWNGGRMEDIPCSRVGHTYRKVYKUPGVSILARN 420

QY 392 RVAEVWMDYEXYTORRPEYRHLASGDYAVOKLRSIANKSKPFRWFMKIAWDLPKF 451

Db 421 RVAEVWMDYEXYTORRPEYRHLASGDYAVOKLRSIANKSKPFRWFMKIAWDLPKF 480

QY 452 PVERPAAWGBEIRNGTGICADTRKGALGSPRLLEGCVGRGRGEAAWNNOVFTWREDI 511

Db 481 PVERPAAWGBEIRNGTGICADTRKGALGSPRLLEGCVGRGRGEAAWNNOVFTWREDI 540

QY 512 RGPDPQHTKCFDFAISHTSPVTLYDCHSMKGNOIWKYRDKTLYHPVSGSCMDCSEDH 571

Db 541 RGPDPQHTKCFDFAISHTSPVTLYDCHSMKGNOIWKYRDKTLYHPVSGSCMDCSEDH 600

QY 572 RIFANTCNSSLTQOWLFENTNSTYLEKEN 601

Db 601 RIFANTCNSSLTQOWLFENTNSTYLEKEN 630

RESULT 4  
US-09-795-926-29

; Sequence 29, Application US/09795926

; Patent No. 655669

; GENERAL INFORMATION:

; APPLICANT: Donoho, Gregory

; APPLICANT: Hilburn, Erin

; APPLICANT: Turner, C. Alexander Jr.

; APPLICANT: Friedrich, Glenn

; APPLICANT: Abuin, Alejandro

; APPLICANT: Zambrowicz, Brian

; APPLICANT: Sands, Arthur T.

; APPLICANT: Walk, D. Wade

; APPLICANT: Wilgowski, Nathaniel L.

; APPLICANT: Hu, Yi

; APPLICANT: Kleke, James Alvin

; APPLICANT: Porter, David George

; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND POLYNUCLEOTIDES ENCODING THE SAME

; FILE REFERENCE: LEX-0144-USA

; CURRENT APPLICATION NUMBER: US/09/795,926

; CURRENT FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: US 60/185,920

; PRIOR FILING DATE: 2000-02-29

; PRIOR APPLICATION NUMBER: US 60/186,558

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 60/191,849

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO: 31

; LENGTH: 535

; TYPE: PRT

; ORGANISM: homo sapiens

; US-09-795-926-29

Query Match 83.8%; Score 2746.5; DB 4; Length 535;

Best Local Similarity 94.6%; Pred. No. 1.; ne-285; Mismatches 0; Indels 29; Gaps 1;

Matches 505; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

QY 158 WSSLRTVWSLVRSPPEVAETVLVDDPSDREHLKKCLEDYMAFFPSVRLTRKREG 217

Db 61 WSSLRTVWSLVRSPPEVAETVLVDDPSDREHLKKCLEDYMAFFPSVRLTRKREG 120

QY 218 IRTMLGASVATGVDITFDSHCEANVNWLPPLDRTARNRKTIVCPMIDVIDDHDFFRYE 277

Db 121 IRTMLGASVATGVDITFDSHCEANVNWLPPLDRTARNRKTIVCPMIDVIDDHDFFRYE 180

QY 278 TOAGDAMGAFDWMETKRPipeLQPELQADSPFPESPVMAGGLFAVDRKWNWELGSDYD 337

Db 181 TOAGDAMGAFDWMETKRPipeLQPELQADSPFPESPVMAGGLFAVDRKWNWELGSDYD 240

QY 338 GLETWGGEOYEISFKWVNGGRMEDIPCSRVGHTYRKVYKUPGVSILARN 397

Db 241 GLETWGGEOYEISFKWVNGGRMEDIPCSRVGHTYRKVYKUPGVSILARN 300

QY 398 MDEAYTIXTORRPEYRHLASGDYAVOKLRSIANKSKPFRWFMKIAWDLPKF 457

Db 301 MDEAYTIXTORRPEYRHLASGDYAVOKLRSIANKSKPFRWFMKIAWDLPKF 360

QY 458 AAMGEIRNGTGICADTRKGALGSPRLLEGCVGRGRGEAAWNNOVFTWREDI 577

Db 421 HTKCFCFDFAISHTSPVTLYDCHSMKGNOIWKYRDKTLYHPVSGSCMDCSEDH 480

QY 518 HTKCFCFDFAISHTSPVTLYDCHSMKGNOIWKYRDKTLYHPVSGSCMDCSEDH 577

Db 421 HTKCFCFDFAISHTSPVTLYDCHSMKGNOIWKYRDKTLYHPVSGSCMDCSEDH 480

QY 578 CNPSSLTQOWLFENTNSTYLEKEN 603

Db 481 CNPSSLTQOWLFENTNSTYLEKEN 506

Query Match 57.3%; Score 1877; DB 4; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-193; Mismatches 0; Indels 0;  
 Matches 353; Conservative 0; MisMatches 0; Gaps 0;

QY 98 MTDADERDOAIRENGFNIVSDKISLNSLSDIRHNCNSKRYLELPTSIIPFHNEG 157  
 1 MTDADERDOAIRENGFNIVSDKISLNSLSDIRHNCNSKRYLELPTSIIPFHNEG 60

QY 158 WSSLURTWISVLNRSPPELVAETIVLWDDFSDREMKKPLEDYMALPPSRLRTRKREG 217  
 61 WSSLURTHSVLSVLNRSPPELVAETIVLWDDFSDREMKKPLEDYMALPPSRLRTRKREG 120

QY 218 IRTMLGASVATDGVITFLSHCRANVNLPLLDIARNRKTVCPMDVIDHDFRYE 277  
 121 IRTMLGASVATDGVITFLSHCRANVNLPLLDIARNRKTVCPMDVITLRTNLSIIPFHNEG 180

QY 278 TQGDAKGAFDNEYMYKRIPPELQKADPSDRESPMAGGLFAVDRKWFELGGYD 337  
 181 TQGDAKGAFDNEYMYKRIPPELQKADPSDRESPMAGGLFAVDRKWFELGGYD 180

QY 338 GLEIWGGEYEISK-----VWMCGRMEDIPCSR 368

Db 241 GLEIWGGEYEISK-----VWMCGRMEDIPCSR 300

QY 369 GHITYKVKPYKVPGVSLARNLKVAEVNDEYAYIYORRPEYRHLASAGDVAVOKKLR 428  
 301 GHITYKVKPYKVPGVSLARNLKVAEVNDEYAYIYORRPEYRHLASAGDVAVOKKLR 360

QY 429 SLNCKSKFWMKTLAWDLKPKYPVPEPPAAGETRNGVGLCAUTKHGAGSPURLEG 488  
 361 SLNCKSKFWMKTLAWDLKPKYPVPEPPAAGETRNGVGLCAUTKHGAGSPURLEG 420

QY 489 VRGRGEAAVNMQFFFWRBDIIGDPOITKCCDAISHTSPTVLICHSMKGQNLWK 548  
 421 VRGRGEAAVNMQFFFWRBDIIGDPOITKCCDAISHTSPTVLICHSMKGQNLWK 480

US-09-795-926-39

Sequence 39, Application US/09795926

QY 549 YRKDKTLYIYRVSQGMDCSBSDHRTFMNTCPNSSTQOMLFENTNSTVLEKFRN 603  
 481 YRKDKTLYIYRVSQGMDCSSESDRHIFMNTCPNSSTQOMLFENTNSTVLEKFRN 535

RESULT 5

US-09-795-926-39

Sequence 27, Application US/09795926

Patent No. 655669

GENERAL INFORMATION:

APPLICANT: Donoho, Gregory

APPLICANT: Hilburn, Brin

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Friedrich, Glenn

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sands, Arthur T.

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Hu, Yi

APPLICANT: Kieke, James Alvin

APPLICANT: Potter, David George

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
 TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795,926

CURRENT FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: US 60/185,920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/185,920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186,558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191,849

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 27

LENGTH: 269

TYPE: PRT

ORGANISM: homo sapiens

US-09-795-926-27

Query Match 41.8%; Score 1370; DB 4; Length 269;  
 Best Local Similarity 100.0%; Pred. No. 9.1e-139; Mismatches 0; Indels 0;  
 Matches 256; Conservative 0; MisMatches 0; Gaps 0;

QY 98 MTDADERDOAIRENGFNIVSDKISLNSLSDIRHNCNSKRYLELPTSIIPFHNEG 157  
 1 MTDADERDOAIRENGFNIVSDKISLNSLSDIRHNCNSKRYLELPTSIIPFHNEG 60

QY 158 WSSLRRTWISVLMNSSPPELVAEVIVLDDFSDRREHKKPLEDYMALFSPVRLRTKRGCL 217  
 Db 61 WSSLRRTWISVLMNSSPPELVAEVIVLDDFSDRREHKKPLEDYMALFSPVRLRTKRGCL 120

QY 218 IRTMLGASVATGIVTFLDSDHCRAVNVLPLPLDRIARNRKTRIVCPMDVHDDFEYE 277  
 Db 121 IRTMLGASVATGIVTFLDSDHCRAVNVLPLPLDRIARNRKTRIVCPMDVHDDFEYE 180

QY 278 TQGDMRGAQDWEMYKIPIPPLQADPSDPESPVMAGGLPAVDRKWFELLGDP 337  
 Db 181 TQGDMRGAQDWEMYKIPIPPLQADPSDPESPVMAGGLPAVDRKWFELLGDP 240

QY 338 GLEIWGGEQEVETISKV 353  
 Db 241 GLEIWGGEQEVETISKV 256

RESULT 7  
 US-09-795-926-35  
 Sequence 35, Application US/09795926  
 Patent No. 655569  
 GENERAL INFORMATION:  
 APPLICANT: Donoho, Gregory  
 APPLICANT: Hibus, Erin  
 APPLICANT: Turner, C. Alexander Jr.  
 APPLICANT: Friedrich, Glenn  
 APPLICANT: Abuin, Alejandro  
 APPLICANT: Zambrowicz, Brian  
 APPLICANT: Sands, Arthur T.  
 APPLICANT: Walke, D. Wade  
 APPLICANT: Wilgowski, Nathaniel L.  
 APPLICANT: Hu, Yi  
 APPLICANT: Kleke, James Alvin  
 APPLICANT: Postter, David George  
 TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
 FILE REFERENCE: LEX-0144-USA  
 CURRENT APPLICATION NUMBER: US/09/795,926  
 CURRENT FILING DATE: 2001-02-28  
 PRIOR APPLICATION NUMBER: US 60/185,920  
 PRIOR FILING DATE: 2000-02-29  
 PRIOR APPLICATION NUMBER: US 60/186,558  
 PRIOR FILING DATE: 2000-03-02  
 PRIOR APPLICATION NUMBER: US 60/191,849  
 NUMBER OF SEQ ID NOS: 47  
 SOFTWARE: Fast-SEQ for Windows Version 4.0  
 SEQ ID NO: 35  
 LENGTH: 321  
 TYPE: PRT  
 ORGANISM: homo sapiens  
 US-09-795-926-35

Query Match 40.0%: Score 1311; DB 4; Length 321;  
 Best Local Similarity 99.2%; Pred. No. 2.8e-132; Mismatches 251; Conservatism 1; Mismatches 1; Indels 0; Gaps 0; Matches 251;

QY 1 MRRKEKRLQAVALVLAIVLPLPGWMLRYERQPDGTGGSAVAPAGGSHSRQK 60  
 Db 1 MRRKEKRLQAVALVLAIVLPLPGWMLRYERQPDGTGGSAVAPAGGSHSRQK 60

QY 61 KTFPLGOGOKKDKMDKEAIRDAQVGNGEQRCPYPMWDAERVQDQYRENGENIYSDK 120  
 Db 61 KTFPLGOGOKKDKMDKEAIRDAQVGNGEQRCPYPMWDAERVQDQYRENGENIYSDK 120

QY 121 ISWRSRSPDPIRPNCSKYLETLPNTSIIPEFNEGHSLLRTVHSYLNRSPELVAEI 180  
 Db 121 ISWRSRSPDPIRPNCSKYLETLPNTSIIPEFNEGHSLLRTVHSYLNRSPELVAEI 180

QY 121 ISWRSRSPDPIRPNCSKYLETLPNTSIIPEFNEGHSLLRTVHSYLNRSPELVAEI 180  
 Db 121 ISWRSRSPDPIRPNCSKYLETLPNTSIIPEFNEGHSLLRTVHSYLNRSPELVAEI 180

QY 181 VLVDFDSREHKKPLEDYMALFSPVRLRTKRGCLTRMLGASVATGIVTFLDSDC 240  
 Db 181 VLVDFDSREHKKPLEDYMALFSPVRLRTKRGCLTRMLGASVATGIVTFLDSDC 240

QY 241 EANVNLPLPLD 253  
 Db 241 EANVNLPLPLD 253

RESULT 8  
 US-08-967-508-19  
 Sequence 19, Application US/08967508  
 Patent No. 5910570  
 GENERAL INFORMATION:  
 APPLICANT: The Upjohn Company  
 APPLICANT: FOR U.S. PURPOSES ONLY: Elhammar, Ake P. and Homa, Fred L.  
 TITLE OF INVENTION: A Cloned DNA Encoding a Up-GaNAc:  
 TITLE OF INVENTION: Polypeptide, N-Acetylgalectosaminyltransferase  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
 ADDRESSEE: Property Legal Services  
 STREET: 301 Henrietta Street  
 CITY: Kalamazoo  
 STATE: Michigan  
 COUNTRY: USA  
 ZIP: 49001  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/967,508  
 FILING DATE: 08/08/967,508  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/602,830  
 FILING DATE: 13 NO. 5910570ember 1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Darley Jr., James D.  
 REGISTRATION NUMBER: 33,673  
 REFERENCE/DOCKET NUMBER: 4755.P CP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 616-833-2210  
 TELEFAX: 616-833-8897  
 TELEX: 224401  
 INFORMATION FOR SEQ ID NO: 19:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 517 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-967-508-19

Query Match 34.0%: Score 1115; DB 2; Length 517;  
 Best Local Similarity 43.6%; Pred. No. 8.3e-11; Mismatches 230; Conservatism 89; Mismatches 170; Indels 38; Gaps 16; Matches 230;

QY 88 GNGQGRP--YPTMDAERVQDQYRENGENIYSDKISLNSPDIRHFNCSKYLETLP 145  
 Db 15 GPGEMGKPKVWPKEDOSKMKEMPINGNFMASEMIANLNSPDIRHFNCSKYLETLP 74

QY 146 NTSLIIPHNEGHSLLRTVHSYLNRSPELVAEVIVLDDFSDRREHKKPLEDYM-ALFP 204  
 Db 75 TISVIVFHNEASTLRTVHSYLNRSPELVAEVIVLDDFSDRREHKKPLEDYM-ALFP 134

QY 205 SVRLRTKRGCLTRMLGASVATGIVTFLDSDHCRAVNVLPLPLDRIARNRKTRIVCP 264  
 Db 135 PVAWRMEQSRGLRARIKGAAVSKGQVITFLDSDHCRAVNVLPLPLDRIARNRKTRIVCP 194

QY 265 MIVDIDHDFPRTQAG-DAMGAFDWEMYKIPPP--ELOKADPSDPESPVMAGG 320  
 Db 195 IIVDIDSDTPEY--MAGSDMTYGGFWKLNFRWYPVQPREMDRKGDRTLPVPTPMAGG 252

QY 321 LFAVDRKWFELLGDPGPLEWCGEQEVISFKWMMCGRMEDIPCSRVRGHIYRKVYPKV 380



STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-02552-19

Query Match 34.0%; Score 1115; DB 5; Length 517;

Best Local Similarity 43.6%; Pred. No. 8.3e-11; Mismatches 170; Indels 38; Gaps 16;

Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGSGGRP--YPMTDAAERVQDQAYRENGFIVYSDKISLRSPLDIRHNCNSKYLETP 145

Db 15 PGGMGKPVIPKEDQEMKEMKINQNLMAISPLPVLEGCKTKYDLP 74

QY 146 NTSLIIPFNGEWSLLRTVHSVANRSPPELVAEVILVUDPSDREHKLKPLLEDY-ALFP 204

Db 75 TISVIVVFNBAWSLRTVHSVINSRSPHMEELIVLVDASERDFLRPLKPLESYVKLV 134

QY 205 SVRLRTKREGGLRTMAGASVATGVDITFLDSHCEANVNMPLRPLRIARKTIVP 264

Db 135 PVRVIRMBORGSLIRARKGAAGVSKGQVITPLDHAECTVWGLPLRARIKHRTVCP 194

QY 265 MIDVJDHDFRYETOQAG-DAMRGADFEMWYKRIPIP--ELQKADSDPFPSPVMAGG 320

Db 195 IIVISDIFEFY--MAGSDMTYGFNMLKANPRTVHSVINSRSPHMEELIVLVDASERDFLRPLKPLESYVKLV 252

QY 321 LFAVDRKWFELGGDGPGLBIRGKQEVYTSFKVNGGSRMEDIPCSRVGHIVYKVPKV 380

Db 253 LFSIDRDFQEGITVDAAGMDIWLGENLEISFRIMOCGGTLEITVCSHVGHVERKATPTP 312

QY 381 PAGVS- -LARNKVAEVYMDDEVAYIYKORRPEYRHLASGDAVOKKURSSLNCKSFWM 418

Db 313 PGGMQQTINNNRNLAEVMDDEKFNPFPYTSPTVKVGDIDSSRLGKHLQCRPSFWY 372

QY 439 MTKIAWD- -LPKFPYPPVEPAANGEIRNGTGLCADTGHAGLSPRLLEGCVGRGEAA 496

Db 422 - -GNQVFSTANKERITDD----LCLDVSKLNGPVTMLKCHLKGNOQLWEYDPVKLTL 473

QY 556 YHPVSGSCMD-CSBDSHRI- -FMTCNCPSSLTQOMLFERTNSTVLEK 600

Db 474 OHVNSNQCLDKATDBDSQVPSIRDGS-SGRSQWLLR- -NVTLPEIF 517

RESULT 11  
US-08-967-508-9

Sequence 9, Application US/08967508

Patent No. 5910570

GENERAL INFORMATION:

APPLICANT: The Upjohn Company

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

APPLICANT: FOR U.S. PURPOSES ONLY: Blahammer, Ake P. and Homa, Fred L.

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,830  
FILING DATE: 13 NO. 5910570ember 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Darinley Jr., James D.  
REGISTRATION NUMBER: 33,673  
REFERENCE/DOCKET NUMBER: 4755.P CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 616-833-2210  
TELEFAX: 616-833-8897  
SEQUENCE CHARACTERISTICS:  
LENGTH: 559 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-967-508-9

Query Match 34.0%; Score 1115; DB 2; Length 559;  
Best Local Similarity 43.6%; Pred. No. 9.5e-11; Mismatches 170; Indels 38; Gaps 16;  
Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGSGGRP--YPMTDAAERVQDQAYRENGFIVYSDKISLRSPLDIRHNCNSKYLETP 145

Db 15 PGGMGKPVIPKEDQEMKEMKINQNLMAISPLPVLEGCKTKYDLP 74

QY 146 NTSLIIPFNGEWSLLRTVHSVANRSPPELVAEVILVUDPSDREHKLKPLLEDY-ALFP 204

Db 177 PVRVIRMBORGSLIRARKGAAGVSKGQVITPLDHAECTVWGLPLRARIKHRTVCP 264

QY 205 SVRLRTKREGGLRTMAGASVATGVDITFLDSHCEANVNMPLRPLRIARKTIVP 264

Db 211 TTSVIVVFNBAWSLRTVHSVINSRSPHMEELIVLVDASERDFLRPLKPLESYVKLV 176

QY 265 MIDVJDHDFRYETOQAG-DAMRGADFEMWYKRIPIP--ELQKADSDPFPSPVMAGG 320

Db 237 IIVISDIFEFY--MAGSDMTYGFNMLKANPRTVHSVINSRSPHMEELIVLVDASERDFLRPLKPLESYVKLV 294

QY 321 LFAVDRKWFELGGDGPGLBIRGKQEVYTSFKVNGGSRMEDIPCSRVGHIVYKVPKV 380

Db 355 LFSIDRDFQEGITVDAAGMDIWLGENLEISFRIMOCGGTLEITVCSHVGHVERKATPTP 354

QY 381 PAGVS- -LARNKVAEVYMDDEVAYIYKORRPEYRHLASGDAVOKKURSSLNCKSFWM 418

Db 395 PGGMQQTINNNRNLAEVMDDEKFNPFPYTSPTVKVGDIDSSRLGKHLQCRPSFWY 372

QY 439 MTKIAWD- -LPKFPYPPVEPAANGEIRNGTGLCADTGHAGLSPRLLEGCVGRGEAA 496

Db 415 LENIYPDQIPIRVY----FSLGERFVNTOCLMARKENEKYGIFNG-HGMG-- 463

QY 497 WNNNQVFTFWRDIPDQDPOHKKFCDAISHTSPVLYDCHSMKGNOQLKRTDK-TL 555

Db 464 - -GNQVFSTANKERITDD----LCLDVSKLNGPVTMLKCHLKGNOQLWEYDPVKLTL 515

QY 556 YHPVSGSCMD-CSBDSHRI- -FMTCNCPSSLTQOMLFERTNSTVLEK 600

Db 516 OHVNSNQCLDKATDBDSQVPSIRDGS-SGRSQWLLR- -NVTLPEIF 559

ADDRESSEE: Property legal Services

STREET: 301 Henrietta Street

STATE: Kalamazoo

COUNTRY: USA

ZIP: 49001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/957,506

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/602,830

FILING DATE: 13 No. 60965/22ember 1995

NAME: Barnley Jr., James D.

REGISTRATION NUMBER: 33,673

REFERENCE/DOCKET NUMBER: 4755.P CP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 616-833-2210

TELEFAX: 616-833-8897

TELEX: 224401

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 559 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLogy: linear

MOLECULE TYPE: protein

US-08-957-506-9

Query Match 34.0%; Score 1115; DB 3; Length 559;

Best Local Similarity 43.6%; Pred. No. 9.5e-111; Mismatches 170; Indels 38; Gaps 16;

Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

Query 88 GNGEOGRP--YPMTDAAERVDQAYRENGFNTYVSDKISLRNLDPIRHPNCNSKRYLELP 145

Db 57 GPGGENGKPVVILPKEDQEKMKFQNKQFLNMASEMIALNRSLDPDRLEGKTKYVDPNL 116

146 NTSLIPFNGEWSLLRTVHSVLAIRSPPELVAETVLVDPDSREHLKPKLEDM-ALFP 204

117 TTSVIVVFINEARWSTLRTVHSVINSRPHMLBIVLDASERDPLKPLIESVKKV 176

205 SVRIRTKKEGLIRTRMLGASVATGDTVLDHCEANVNWLPPLDTRARNRTIVCP 264

177 FVHVIRMEQSGLIRLARLGAVSNGQVITFLDACECTVGLQLEPLALARLKHDKTVC 236

Db 265 MIDVIDHDDPRYEQOAG-DAMRGAFDPMWEMYKRPPIP--ELQKADSDPPSPVMAG 320

Db 237 IDIVSDDTPPEY--MAGSDMTYGGFNRWKLMRWRVYVPPOREMDRKKDRTRPVRTPMAGG 294

321 LFAVDRKWFELGGDGPGLERWGGQYESFVKWMGGRMEDIPCSRVRGHYRKYVQPYKV 380

Db 295 LFSIDRDYFOBIGTYDAGMDIWGGENLEISPRIWCGGLELIVTCSHVGHVFRKATPYF 354

381 PAGVS--LARNLKRVAVWMDDEAYIYKORPPEYRHLISAGDVAVQKKLRSLLNCSPKWF 438

355 PGGTQGILINKNRRRLAEWMDDEFKNPFLYIISPGVTKYDQDSSNLGRKLUQRPFSWY 414

Db 439 NTKIAND--LPKFVPPVEPPAAWGSFIRNTYGTGLCADTRKGALGSPSLRGGCVRGEBAA 496

Db 415 LENIVFDSQFPRHY----PSLGEIRNBTENOCLDNMARKENKEVKGIFNC-HOMG-- 463

497 WNNMVOFTTFRDTRPGDPQHTKCFDAISHTSPVTLYDCHSMKGNOIWKYRKDK-TL 555

Db 464 --GNQFSVYNTANKEIERTD----LCLDVSKLNGPVTMLKCHHLKGNOIWEYDPVKLTL 515

556 YHPVSGSCMD-CSESHPRI FMTCOPSSLTQQMFLFENTSTIVLKF 600

Db 516 QHVNNSNQCLDQKATEDSQVPSIRDCS-GSSRSQWLLR--NVTLRPEIF 559

RESULT 13

PCN-US94-02552-9

Sequence 9, Application PC/TUS9402552

GENERAL INFORMATION:

APPLICANT: Blahammer, Ake P.

APPLICANT: Homa, Fred L.

TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAc:

TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: The Upjohn Company, Corp. Intellectual

STREET: 301 Henrietta Street

CITY: Kalamazoo

STATE: Michigan

COUNTRY: USA

ZIP: 49001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PC/TUS94/02552

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Barnley Jr., James D.

REGISTRATION NUMBER: 33,673

REFERENCE/DOCKET NUMBER: 33,4755.P CP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 616-385-5210

TELEFAX: 616-385-6897

TELEX: 224401

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 559 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLogy: linear

MOLECULE TYPE: protein

PCT-US94-02552-9

Query Match 34.0%; Score 1115; DB 5; Length 559;

Best Local Similarity 43.6%; Pred. No. 9.5e-111; Mismatches 170; Indels 38; Gaps 16;

Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

Query 88 GNGEOGRP--YPMTDAAERVDQAYRENGFNTYVSDKISLRNLDPIRHPNCNSKRYLELP 145

Db 57 GPGGENGKPVVILPKEDQEKMKFQNKQFLNMASEMIALNRSLDPDRLEGKTKYVDPNL 116

146 NTSLIPFNGEWSLLRTVHSVLAIRSPPELVAETVLVDPDSREHLKPKLEDM-ALFP 204

117 TTSVIVVFINEARWSTLRTVHSVINSRPHMLBIVLDASERDPLKPLIESVKKV 176

205 SVRIRTKKEGLIRTRMLGASVATGDTVLDHCEANVNWLPPLDTRARNRTIVCP 264

177 FVHVIRMEQSGLIRLARLGAVSNGQVITFLDACECTVGLQLEPLALARLKHDKTVC 236

Db 265 MIDVIDHDDPRYEQOAG-DAMRGAFDPMWEMYKRPPIP--ELQKADSDPPSPVMAG 320

Db 237 IDIVSDDTPPEY--MAGSDMTYGGFNRWKLMRWRVYVPPOREMDRKKDRTRPVRTPMAGG 294

321 LFAVDRKWFELGGDGPGLERWGGQYESFVKWMGGRMEDIPCSRVRGHYRKYVQPYKV 380

Db 295 LFSIDRDYFOBIGTYDAGMDIWGGENLEISPRIWCGGLELIVTCSHVGHVFRKATPYF 354

381 PAGVS--LARNLKRVAVWMDDEAYIYKORPPEYRHLISAGDVAVQKKLRSLLNCSPKWF 438

355 PGGTQGILINKNRRRLAEWMDDEFKNPFLYIISPGVTKYDQDSSNLGRKLUQRPFSWY 414

Db 439 NTKIAND--LPKFVPPVEPPAAWGSFIRNTYGTGLCADTRKGALGSPSLRGGCVRGEBAA 496

Db 415 LENIVFDSQFPRHY----PSLGEIRNBTENOCLDNMARKENKEVKGIFNC-HOMG-- 463

497 WNNMVOFTTFRDTRPGDPQHTKCFDAISHTSPVTLYDCHSMKGNOIWKYRKDK-TL 555

Db 464 --GNQFSVYNTANKEIERTD----LCLDVSKLNGPVTMLKCHHLKGNOIWEYDPVKLTL 515

556 YHPVSGSCMD-CSESHPRI FMTCOPSSLTQQMFLFENTSTIVLKF 600

Db 321 LFAVDRKWFELGGDGPGLERWGGQYESFVKWMGGRMEDIPCSRVRGHYRKYVQPYKV 380

Db 295 LFSIDRDYFOBIGTYDAGMDIWGGENLEISPRIWCGGLELIVTCSHVGHVFRKATPYF 354

381 PAGVS--LARNLKRVAVWMDDEAYIYKORPPEYRHLISAGDVAVQKKLRSLLNCSPKWF 438

355 PGGTQGILINKNRRRLAEWMDDEFKNPFLYIISPGVTKYDQDSSNLGRKLUQRPFSWY 414

RESULT 14  
US-09-217-306B-2  
; Sequence 2, Application US/09217306B  
; GENERAL INFORMATION:  
; PATENT NO. 6465220  
; APPLICANT: Hassen, Helle  
; APPLICANT: Clausen, Henrik  
; APPLICANT: Bennett, Eric P.  
; TITLE OF INVENTION: Glycosylation Using GalNAc-T4 Transferase  
; FILE REFERENCE: 8850\*1  
; CURRENT APPLICATION NUMBER: US/09/217,306B  
; CURRENT FILING DATE: 1998-12-21  
; CURRENT FILING NUMBER: US/09/217,306B  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 578  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-217-306B-2

Query Match 31.3%; Score 1026; DB 4; Length 578;  
Best Local Similarity 39.4%; Pred. No. 3\_9e-101; Indels 38; Gaps 16;  
Matches 221; Conservative 98; Mismatches 204; SEQ ID NO 10

Qy 50 AAGCGSHSKOKTKTRPLGDEQKTKWHDKEAARR--DAQPRVGEQGRGPYPMDAE--RV 104  
Db 35 SAGAG-RARBLGSRRLSDQKNTEDLSRPLKKEPPADSRALGENGKASKLQLNEDLKQ 93

Qy 105 DQAYRENGNIVYSDKISLARSLPDIRHNCNSKRY-LETLPLNTSIIIPFHNEGWSLIR 163  
Db 94 ELLERYAINTYLSDRISLHRSDKRMTECKSQKENTYPLTISVIAFYNEAWSTLIR 153

Qy 164 TUVSYLNRSPPELVAEVIVDSDREHKKPLDYMALFSSVYIIRTKRREGIIRTL 223  
Db 154 THSVLETSPAVLKEIILVDDSDRVYLTQLETYISMDRVLRILTRKREGI 213

Qy 224 GASVATGIVTFLSHCREANVWNPPLLARIARRKITWCPIVQDIDRPRVETQADA 283  
Db 214 GATFATGIVTFLYCHCECNCGWLEPLLERGRYETAVVCPVIDTIDNTEFVQICP 273

Qy 284 MGRGDWEMVYKRIPIPPEL--QKADSPDPFESPVMAGGLPAVDRKWFELGGDGPGLI 341  
Db 274 MIGERDWRITFQWHSVPKQERDRISRIPRSPTRSPMAGLPAVSKYKRYQYLGTVIG 333

Qy 342 WGGROYETSPFKVNGGGRDIDPESRVRGHYKRYKPYKPGAVSLARNKRVARVWMDY 401  
Db 334 WGGENLELSFRVWQEGGKUBIHPCHVHGKVFPPKQAPYAP--NFLONTARAKLWMDY 390

Qy 402 AEVYQRRPRYRHSAGDVAVWOKLRSIUNCKSTKWMFKIWLPLKFVPPVERPAAM- 460  
Db 391 KEHFTYRNRPARKERAKYGDISERKULRERLCKSFDWYKLWV--FPNLHVPEDP--GWH 445

Qy 461 GEIRAVG-TGLCAD--TRIGALGSPLRLEGCVGRGEAMNNMQVFTWREDIRPGD 516  
Db 446 GATRSRGISCELDNTPNMTGANTSLJFGC-HGQG---GNOFFETNSKEIRFN- 498

Qy 517 OHTKFKFCFDAISHTSPVTLDC---HSMKGNOIWKYRDKTLYHIVPSSSCMDC---SBS 569  
Db 499 -VTELCAEVPEOKNYVMQNCPKDGFPPVPAINTWHFKEDGTIFPHSGSLCLSAVTPCG 556

Qy 570 DHRIFMNTCPSSLTQOPLF 590  
Db 557 RPDVQMRCTDALKNQIWSFE 577

RESULT 15  
US-09-217-306B-10  
; Sequence 10, Application US/09217306B  
; Patent No. 6465220  
; GENERAL INFORMATION:  
; APPLICANT: Hassen, Helle  
; APPLICANT: Clausen, Henrik  
; APPLICANT: Bennett, Eric P.  
; TITLE OF INVENTION: Glycosylation Using GalNAc-T4 Transferase  
; FILE REFERENCE: 8850\*1  
; CURRENT APPLICATION NUMBER: US/09/217,306B  
; CURRENT FILING DATE: 1998-12-21  
; NUMBER OF SEQ ID NOS: 25  
; SEQ ID NO 10  
; LENGTH: 560  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Interferon fragment-GalNAc-T4 predicted coding region  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(25)  
; OTHER INFORMATION: human interferon signal peptide  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (37)..(560)  
; OTHER INFORMATION: (37)..(560)  
; OTHER INFORMATION: GalNAc-T4  
; US-09-217-306B-10

Query Match 31.2%; Score 1024; DB 4; Length 560;  
Best Local Similarity 41.0%; Pred. No. 6.1e-101; Indels 36; Gaps 14;  
Matches 213; Conservative 88; Mismatches 183; SEQ ID NO 11

Qy 90 GEOPRYPM---TDAERVDQAYRENGNIVYSDKISLARSLDIRHNCNSKRY-LETL 144  
Db 57 GEWGRASKQIQLNEDLKQELBLKQELIERYAINTYLSDRISLHRSDKRMTECKSKOFNTL 116

Qy 145 PNTSIIIPFHNEGWSLIRTKRREGIIRTL 223  
Db 145 PNTSIIIPFHNEGWSLIRTKRREGIIRTL 223

Qy 177 RVLIRTKRREGIIRLGRIGATATGDPVLFUCHCECNCGWLEPLLERGRYETAVVCP 236  
Db 177 RVLIRTKRREGIIRLGRIGATATGDPVLFUCHCECNCGWLEPLLERGRYETAVVCP 236

Qy 205 SVRLIRTKRREGIIRLGRIGATATGDPVLFUCHCECNCGWLEPLLERGRYETAVVCP 264  
Db 205 SVRLIRTKRREGIIRLGRIGATATGDPVLFUCHCECNCGWLEPLLERGRYETAVVCP 264

Qy 265 MIDVHDHPFRTYDAGAAMGRGDWEMVYKRIPIPPEL--QKADSPDPFESPVMAGGLP 322  
Db 237 VIDTIDWNTFEPFMQIGPMIGEDWRITFQWHSVPKQERDRISRIPRSPFMAGGLP 296

Qy 323 AUPRKWFELGGYDPLGLIWGGQYELSSPKVNGGGRDIDPESRVRGHYKRYKPYKPA 382  
Db 323 AUPRKWFELGGYDPLGLIWGGQYELSSPKVNGGGRDIDPESRVRGHYKRYKPYKPA 382

Qy 297 AVSKYKFCYQYLGTVDTGMETWGGNTELELSFRVWQCGGKUBIHPCHVHGKVFPPKQAPYAP- 355

Qy 383 GVSGLARNKRVAVWMDYEAEVYQRRPRYRHSAGDVAVWOKLRSIUNCKSTKWMFKI 442  
Db 356 --NFLQNTARAKYGDISERKULRERLCKSFDWYKLWV--FPNLHVPEDP--GWH 413

Qy 443 AWDLPLKPTPVEPPAAN-GEIRAVG-TGLCAD--TRIGALGSPLRLEGCVGRGEAMN 497

Db 414 --FPNLHVPEDP--GNGAATRSRGISCELDNTPNMTGANTSLJFGC-HGQG--- 463

Qy 498 NNMQVFTTWRDPRPGCQPTKFCFAISHTSPVTLDC---HSMKGNOIWKYRDKTLYHIVPSSSCMDC---SBS 553  
Db 464 -GNOFFETNSKEIRFN- --VTELCAEVPEOKNYVMQNCPKDGFPPVPAINTWHFKEDGTIFPHSGSLCLSAVTPCG 519

Qy 554 TLYHVSQSCMDC---SSSDHRIFMNTCPSSLTQOPLF 590

RESULT 16  
 Sequence 33, Application US/09795926  
 Patent No. 6555669  
 GENERAL INFORMATION:  
 APPLICANT: Donoho, Gregory  
 APPLICANT: Hilbun, Brian  
 APPLICANT: Turner, C. Alexander Jr.  
 APPLICANT: Friedrich, Glenn  
 APPLICANT: Abuin, Alejandro  
 APPLICANT: Sands, Arthur T.  
 APPLICANT: Walk, D. Wade  
 APPLICANT: Wilgnowski, Nathaniel L.  
 APPLICANT: Hu, Yi  
 APPLICANT: Kieke, James Alvin  
 APPLICANT: Potter, David George  
 TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
 TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
 FILE REFERENCE: LEX-0144 USA  
 CURRENT APPLICATION NUMBER: US/09795926  
 CURRENT FILING DATE: 2001-02-28  
 PRIOR APPLICATION NUMBER: US 60/185,920  
 PRIOR FILING DATE: 2000-03-29  
 PRIOR APPLICATION NUMBER: US 60/186,558  
 PRIOR FILING DATE: 2000-03-24  
 PRIOR APPLICATION NUMBER: US 60/191,849  
 PRIOR FILING DATE: 2000-03-24  
 NUMBER OF SEQ ID NOS: 47  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO: 37  
 LENGTH: 209  
 TYPE: PRT  
 ORGANISM: homo sapiens  
 US-09795926-33

Query Match 30.2%; Score 990; DB 4; Length 240;  
 Best Local Similarity 97.9%; Pred. No. 6.8e-98; Matches 189; Conservative 2; Mismatches 2; Indels 0; Gaps 0; Gaps 0;  
 Matches 189; Conservative 2; Mismatches 2; Indels 0; Gaps 0; Gaps 0;

QY 1 MRRKEKRLQVALVALVAVLPNGWALRERQDGTCGGSGAVAPAGQSHSRQK 60  
 Db 1 MRRKEKRLQVALVALVAVLPNGWALRERQDGTCGGSGAVAPAGQSHSRQK 60

QY 61 KTFPLGDGQKLDKMDKEAIRRDAQRGVNGCGRPYPMTDERVDAQYRENGFNIVSDK 120  
 Db 61 KTFPLGDGQKLDKMDKEAIRRDAQRGVNGCGRPYPMTDERVDAQYRENGFNIVSDK 120

QY 121 ISLNSLSPDRHPCNSKRYLETLPNTSIIIPFHNEGSSLRVTVHSVLRSPPELVASI 180  
 Db 121 ISLNSLSPDRHPCNSKRYLETLPNTSIIIPFHNEGSSLRVTVHSVLRSPPELVASI 180

QY 181 VLVDDPSDR 189  
 Db 181 VLVDDPSDR 189

RESULT 17  
 US-09795926-37  
 Sequence 37, Application US/09795926  
 Patent No. 6555669  
 GENERAL INFORMATION:  
 APPLICANT: Donoho, Gregory  
 APPLICANT: Hilbun, Brian  
 APPLICANT: Turner, C. Alexander Jr.  
 APPLICANT: Friedrich, Glenn  
 APPLICANT: Abuin, Alejandro  
 APPLICANT: Sands, Arthur T.  
 APPLICANT: Walk, D. Wade  
 APPLICANT: Wilgnowski, Nathaniel L.  
 APPLICANT: Hu, Yi  
 APPLICANT: Kieke, James Alvin  
 APPLICANT: Potter, David George  
 TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
 TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
 FILE REFERENCE: LEX-0144 USA  
 CURRENT APPLICATION NUMBER: US/09795926  
 CURRENT FILING DATE: 2001-02-28  
 PRIOR APPLICATION NUMBER: US 60/185,920  
 PRIOR FILING DATE: 2000-03-28  
 PRIOR APPLICATION NUMBER: US 60/191,849  
 PRIOR FILING DATE: 2000-03-24  
 NUMBER OF SEQ ID NOS: 47  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO: 37  
 LENGTH: 209  
 TYPE: PRT  
 ORGANISM: homo sapiens  
 US-09795926-37

Query Match 30.1%; Score 988; DB 4; Length 209;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-98; Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0; Gaps 0;

QY 1 MRRKEKRLQVALVALVAVLPNGWALRERQDGTCGGSGAVAPAGQSHSRQK 60  
 Db 1 MRRKEKRLQVALVALVAVLPNGWALRERQDGTCGGSGAVAPAGQSHSRQK 60

QY 61 KTFPLGDGQKLDKMDKEAIRRDAQRGVNGCGRPYPMTDERVDAQYRENGFNIVSDK 120  
 Db 61 KTFPLGDGQKLDKMDKEAIRRDAQRGVNGCGRPYPMTDERVDAQYRENGFNIVSDK 120

QY 121 ISLNSLSPDRHPCNSKRYLETLPNTSIIIPFHNEGSSLRVTVHSVLRSPPELVASI 180  
 Db 121 ISLNSLSPDRHPCNSKRYLETLPNTSIIIPFHNEGSSLRVTVHSVLRSPPELVASI 180

QY 181 VLVDDPSDR 189  
 Db 181 VLVDDPSDR 189

RESULT 18  
 US-10140-002-196  
 Sequence 196, Application US/10140002  
 Patent No. 6725750  
 GENERAL INFORMATION:  
 APPLICANT: Baker, Kevin P.  
 APPLICANT: Bersini, Maureen  
 APPLICANT: DeForge, Laura  
 APPLICANT: Destroyer, Luc  
 APPLICANT: Flavoff, Ellen  
 APPLICANT: Gao, Wei-Qiang  
 APPLICANT: Gerritsen, Mary B.  
 APPLICANT: Goddard, Audrey  
 APPLICANT: Godowski, Paul J.  
 APPLICANT: Gurney, Austin L.  
 APPLICANT: Sherwood, Steven  
 APPLICANT: Smith, Victoria  
 APPLICANT: Stewart, Timothy A.  
 APPLICANT: Tumab, Daniel  
 APPLICANT: Watanabe, Colin K  
 APPLICANT: Wood, William

APPLICANT: Zhang, Zemin  
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEARIC  
 TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 FILE REFERENCE: P3330R1C59  
 CURRENT APPLICATION NUMBER: US/10140,002  
 CURRENT FILING DATE: 2002-05-06  
 Prior Application removed - See Palm or File Wrapper



QY 358 GRMEDIPCSPTVGHIVKRYVVKPAGVSLARNLKEVAEVMDEAVYIYQRPERVHLA 417  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 405 GSVEILPCSRVGHIVYQNDOSHSPDQEATURNRRAETWIGSFETFKHSPFPLSK 464  
 QY 418 G--DVAVKRLRSINLCKSFKWFNTKIAWDLRKFVPPVEPAAWGRIRNVGTLGCA- 473  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 465 AEKPDCMERLQLQRLRGCRTHFWFLANV--YPELYPS-BPRPSFSGKLNHTGLGCA 520  
 QY 474 -TKIGALGSPLRLCVRGRGEAAWNMQVFTWREDRGGDQHTKCFDASHTSP 520  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 521 QAEQDILGCPMVLAPCSISR-----QOYLOHTSRKIHFGSPQH--LCFAV--RQHQ 569  
 521 QAEQDILGCPMVLAPCSISR-----QOYLOHTSRKIHFGSPQH--LCFAV--RQHQ 569  
 QY 533 VTLVDCHSMKG---NOLWKRKDKTLYHPVSGSCMDC--SESDHRIMNTCNSSLTOQ 586  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 570 VILQNC-TBGLATHQHQDFOENGMINHILSGKOMEAVVQENNDLILRCPD-GKARQ 627  
 QY 587 WLFEHTNS 594  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 628 WRFDOINA 635

RESULT 21  
 Sequence 2, Application US/09347488  
 Patent No. 6239466  
 GENERAL INFORMATION:  
 APPLICANT: Monroe, David  
 APPLICANT: Pribill, Ingrid  
 TITLE OF INVENTION: USES  
 TITLE OF INVENTION: USES  
 FILE REFERENCE: GEN-2PRV  
 CURRENT APPLICATION NUMBER: US/09/347,488  
 CURRENT FILING DATE: 1999-07-02  
 NUMBER OF SEQ ID. NOS: 2  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 2  
 LENGTH: 638  
 TYPE: PRT  
 ORGANISM: H. sapiens  
 US-09-347-488-2

Query Match 28.1%; Score 920; DB 3; Length 638;  
 Best Local Similarity 35.4%; Pred. No. 1.2e-89;  
 Matches 215; Conservative 96; Mismatches 217; Indels 80; Gaps 17;

QY 4 KERKLQAVAVLVALVLLPVLNGLWALYRERQDGTGGGAAVAPACGGSHSRQKKT 63  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 91 REDOILWAVALPOA-----RINOSCCRGGSYRLIKQPRRQ----- 126  
 91 REDOILWAVALPOA-----RINOSCCRGGSYRLIKQPRRQ----- 126  
 QY 64 FLGDQGQLKUDWHDKEAIRRDQARVNGEQRYPYMT---DAERVDQAYRENGNIYSD 119  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 127 -----DKSAPKRDGADEGEVSEEEETPPSLDPRCQEA-----USA 165  
 127 -----DKSAPKRDGADEGEVSEEEETPPSLDPRCQEA-----USA 165  
 QY 120 KISLNRSLPDRHPRCNSKRYLETPLNTSITIPFNEGNSSLRIVHSVNRSPPELV 179  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 166 RIPLORALPVRHPLCQHQDLSPTASVILCFHDWBASILLRTVHSILDTVRAFLKE 225  
 166 RIPLORALPVRHPLCQHQDLSPTASVILCFHDWBASILLRTVHSILDTVRAFLKE 225  
 QY 180 IVLVDPSDREHHLKPLEDTWALFSVTRTKEGLIRTRMIGASVATGQVITPLDHS 239  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 226 ITLVDLSSQQQLKSLASBYVARLBGVKLRSKRSIARMLGATRANGDVLVMDAH 285  
 226 ITLVDLSSQQQLKSLASBYVARLBGVKLRSKRSIARMLGATRANGDVLVMDAH 285  
 QY 240 CEAANWLPILDLRARNRKTIVCMIDVHDHDDREYETQGAMGAFDWEYMYKRIPI 299  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 286 CECHPGWLEDLISRSRAGDRRSVSPVSDIVDVKWTOY-YSSKDLORGVDLWPHWEPL 344  
 300 PPELOQKA--DPSDPRPSPVMGGFLPDRKPNWFLWGLGDDGELTWGEOYEISPKWMC 357  
 345 PEHVRKALQSPDISPISPRSPVVEGVVAMDRHIVFQNTGAYDLSMSLKGGENLELSFKAWLCG 404  
 358 GRMEDIPCSRKGHTIRKVKVPGVSLARNLKVAEVMDVEYIYORRPERVHLA 417  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 405 GSVEILPCSRVGHIVYQNDOSHSPDQEATURNRRAETWIGSFETFKHSPFPLSK 464

RESULT 22  
 US-08-648-298-2  
 Sequence 2, Application US/08648298  
 Patent No. 5875990  
 GENERAL INFORMATION:  
 APPLICANT: Henrik Clausen  
 APPLICANT: Eric Paul Bennett  
 TITLE OF INVENTION: UDP-N-acetyl- $\alpha$ -D-galactosamine:polypeptide  
 TITLE OF INVENTION: N-acetylgalactosaminyltransferase Gainac-T3  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Darby & Darby PC  
 STREET: 805 Third Avenue  
 CITY: New York  
 STATE: NY  
 ZIP: 10022  
 COMPUTER READABLE FORM:  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: Patientin Release #1.0, Version #1.30 (BPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/648,298  
 FILING DATE: 15-JUN-1996  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Green, Rose  
 REGISTRATION NUMBER: 38,475  
 REFERENCE/DOCKET NUMBER: 4035/0B865  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 21227700  
 TELEFAX: 2127536237  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 633 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 TISSUE TYPE: Submaxillary gland  
 US-08-648-298-2

Query Match 27.9%; Score 916; DB 2; Length 633;  
 Best Local Similarity 37.6%; Pred. No. 3.3e-89;  
 Matches 202; Conservative 100; Mismatches 177; Indels 58; Gaps 18;

QY 90 GEOGRYPWPMDAERTDAVRENG---FNIYVDSKRSIARSL\_PDIRUPNCNSRY--L 141  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 122 GASGKAKTINLNSVBEQKKGKRGKCFNCFNAFASDRISLHDLGPDRPPECQFKRC 181  
 142 ETLPNTSITPFHNGEGWSSLURTHSVNRSPELVAEVVLVUDPSDRHHLKKELEDNA 201  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 182 PPLPTMSVIVFHNBAWSTLRTVHSVLYSSPAILKELILVDDASVBDYLHDKLDYVK 241

Qy 202 LFFSPVRLRTKREGLIRTRMLGASVATGVDVTFDSHCEANVNLPLPLRARNKTI 261  
 Db 242 QFSIVKIVRQRGLTARLGLGATVATBTLTFLDAHCEFCYCWLEPLIARLNTAYAV 301

Qy 262 VCPMDIVDHDPRYETOA--GDMRGAEDWENY--KRIPPFLQKADRDPEPRV 316  
 Db 302 VSPDIASIDINTTFBNKESPYGSNHRGNFDWSLSPGMSLSPDHEKQRKDPRYPIKPT 361

Qy 317 MAGGLFAVDRKWFELGGDPPGLIWIWGGQYELSPFKVMCGRMEDIPCSRVRGHYRKV 376  
 Db 352 PAGGLFSISKEFVYIGSDEMEWGGMINERSPRVWQCGGLEIMCSPVGVPRKS 421

Qy 377 PYKVPGAVS-LARMLKRVLEWVWMDDEAYTIVRRPE---YRHLUSAGBVAVKLRSLN 431  
 Db 422 PHSPPKGTOVIAARNQVRLAEWMDBYKEFVYRNRVDAAKIVKOKAFGDLISKRFIKHLR 481

Qy 432 CKSKWPKMVKLAWLPKFEVPPVERPAAAGELRVRGTLGACT-KHAGLSPLEGUR 490  
 Db 482 CKNTFTWYLNNI--YPEVTPDLPVIS-GYKNSVGQPLCUDVGENNODGKPLIMTC-H 536

Qy 491 GRGEAAWNNMQVFTWREDRIPGDQHT--KKKCFDAT---SHSPVTLY 536  
 Db 537 GLG---GNOVYSAQHRI---HNQKELCHHAQGLVOLKACTYKGKIVT-- 594

Qy 537 DCHSMKGNOIWKYRKDTLYHPVSGSCMCSESCHRIFANTCPSSTOOWLFEHTN 593  
 Db 585 -----GEOWEIORDQDOLLYNPLKJNCUS-ANGEHPSLV-SCNPSDPLQKWLSOND 633

RESULT 23  
 US-09-975-926-23  
 Sequence 23, Application US/0975926  
 Patent No. 655569

GENERAL INFORMATION:

APPLICANT: Donoho, Gregory  
 APPLICANT: Hilbun, Brin  
 APPLICANT: Turner, C. Alexander Jr.  
 APPLICANT: Friedrich, Glenn  
 APPLICANT: Abuin, Alejandro  
 APPLICANT: Zambrowicz, Brian  
 APPLICANT: Sands, Arthur T.  
 APPLICANT: Walko, D. Wade  
 APPLICANT: Wilgowski, Nathaniel L.  
 APPLICANT: Hu, Yi  
 APPLICANT: Kile, James Alvin.

APPLICANT: Potter, David George

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
 TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/975, 926

CURRENT FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: US 60/185, 920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186, 558

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 23  
 LENGTH: 224  
 TYPE: PRT  
 ORGANISM: homo sapiens

Query Match 22.7% Score 744; DB 4; Length 603;  
 Best Local Similarity 32.3%; Pred. No. 1e-70; Gaps 23;  
 Matches 202; Conservative 86; Mismatches 230; Indels 108; Gaps 23;

Qy 3 RKEKRLQAVAUALLLPNPLNGLWYRERODPDTGGSGAVAPAG-QOSSRKK 61  
 Db 5 RKRTLTVNLVLFVGV-----LFSYTCRQ-----GRSOLVRLTVSGDRVRSHAK 53

Qy 62 TFFLGDGOKKDW--HDEKAIARRDAQRV-----GNGRQGRPYM-TDAERVQOA 107  
 Db 54 VGTIGDRAILQDHLHRSWVNLQNGLAAPKIGLVEGGGIGGCGGLATLROGQEAQGK 113

Qy 108 YRENGFTNTVSKUSLNSTLDPDRHNPNSKRYLTENNTS1PFHFGWSSLRTHS 167  
 Db 114 YEBGYNQASUDRISLSDSIPDTRPKRQMSAQDQVLSVVFIFNALSVLRSVS 173

Qy 168 VLRSPPLVAAEVLDVDSRHLKLEDM-ALPS-VRLRTKREGLIRTRMGA 225  
 Db 174 VVNRTPSOLKEVILVLDNSDNTVKEVLDLARLNTDOVKNVRYGLVKTVRNSREGLRTARLQGW 233

Qy 226 SVAIGDVTFLDSHCEANVNLPLPLDILARLNTKTIQPMIDVHDHPDPRYETOQADMVR 285  
 Db 234 KATAAPVUGFDPAVEFTGMAWPAFLSIREBRRIVPAIMIKYSTFEVO-QYANAH 292

Qy 286 GAFDWMEMYKRIIPPE--LOKADPSDFESPVMAGGFAVFKWFNLGGDPGLEW 343  
 Db 293 G-YNWLGLNOMYL-IPOWLDRGDESAP-IRTPAMIGESFVVDREYFGDIDLUDPGMEVY 350

Qy 344 GGEVBLSKWMMCGRMEDIPCSRVRGHYRKVYKPVAGVSLARNKRVAVTMDEXA 403  
 Db 351 GENVELGRVWOGGSMVTPESRVAHIERTRKPYNNIDYAKRNARAAEVYMDDEKS 410

Qy 404 YIVQR--RPEYRHLISAGDVAVOKKRSLSINCKSFKRMVKLAWDLPKFYP--VEPPAA 458  
 Db 411 HVTMAMNTPMSNGCVDGDSVSLALRQLKESFKYMLNT-----YPEVYNTL 463

Qy 459 AWERBN VGTGLCADTHKAGLSPRLLEGCVGRGSEAWNMQVFTWREDRIPGPO 517  
 Db 464 TYGEVRNSKASAVCLD--OGA-----ED--GD-- 486

Qy 98 MTDAAERVDOAYRENGFNTVSKDKISLNRSLDPDRHPCNSKRYLETPLNTSITIPFREG 157  
 Db 1 MTDAAERVDOAYRENGFNTVSKDKISLNRSLDPDRHPCNSKRYLETPLNTSITIPFREG 60

Query Match 24.5% Score 804; DB 4; Length 224;  
 Best Local Similarity 98.7%; Pred. No. 6.7e-78; Matches 154; Mismatches 1; Indels 0; Gaps 0;

```

Db 487 -----RAILYPCGHM-SSQULVYSAQGLQPLGSTAFPLDS-KCLVDDG 530
Qy 569 SDHRIFMNTCPSSUTQWLFENTNS 594
Db 531 TGRMPTLKKCEDVARPTQRQLWLDFTOS 556

RESULT 25
US-09-270-767-45334
; Sequence 45334, Application US/09270767
; Patent No. 670391
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45334
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-45334

```

```

Query Match 20.6%: Score 675; DB 4; Length 289;
Best Local Similarity 50.2%; Pred. No. 8.2e-64; Mismatches 77; Indels 8; Gaps 4;
Matches 131; Conservative 45; MisMatches 77; Indels 8; Gaps 4;

Qy 111 NGPNIVPSDKISLMSLSPDPRHPCNGNSKRYLETLNTSIIPIFNEGWSLRLTVHSVNL 170
Db 34 NRENQEAASDAPSLPSNRDIPDTRNPMDRCRTKYYKREDBETPSVTTNEARSTPLRTIVSVLN 93
Qy 171 RSPPELVAETVLVDDFSDRHLLKKCLEDYMAFLFSPVRLKTKRREGLIRTRMLGASVATG 230
Db 94 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 149
Db 231 DVITFLDSKCEANVWLPPLDRIANRKTIVCPMIDVDDDFRYETORGDAWGAFW 290
Qy 150 SVLTFLDSHVCENEWLPEPLIERVERVEDPTAVCPVTPIDVSDWDFNFOVIGASAD-LRGGFDW 208
Qy 291 EMVYKRPPIPPE--LQKADPSDPFESPVMAGGLEAVDRKFWELGGYGPGLETWGGEOY 347
Db 209 NLIFKWEVYLSPSERAMRHNDPTTAIRTPMIAGLFVIDKAYENKLKGKDMVWGENL 268
Qy 348 EISPKWWMCGGERMEDIPCSV 368
Db 269 EISFRVWQCGGSLIEIPCSV 289

```

Search completed: November 22, 2004, 13:44:43  
 Job time : 23 sec(s)

GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

Om protein - protein search, using sw model  
 Run on: November 22, 2004, 13:37:52 ; Search time 57 Seconds  
 (without alignments)  
 3746.297 Million cell updates/sec

Title: US-10-001-851-2  
 Perfect Score: 3278  
 Sequence: 1 MRKKERKRLGCAVALVLAALV.....TQOWLFEHTNSTVLRKFNRN 603

Scoring table: BLOSUM62

Gapop 10.0 , Gapext: 0.5

Searched:

1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0  
 Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
 Listing first 100 summaries

Database : Published Applications At: \*  
 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep: \*  
 2: /cgn2\_6/ptodata/1/pubpaa/PCPT\_NEW\_PUB.pep: \*  
 3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep: \*  
 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep: \*  
 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep: \*  
 6: /cgn2\_6/ptodata/1/pubpaa/PCUTS\_PUBCOMB.pep: \*  
 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep: \*  
 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep: \*  
 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep: \*  
 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep: \*  
 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep: \*  
 12: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep: \*  
 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep: \*  
 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep: \*  
 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep: \*  
 16: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep: \*  
 17: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep: \*  
 18: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep: \*  
 19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep: \*  
 20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	3278	100.0	603	9 US-09-795-926-43	Sequence 43, Appl
2	3278	100.0	603	13 US-10-001-851-2	Sequence 2, Appl
3	3278	100.0	603	14 US-10-364-774-43	Sequence 43, Appl
4	3242.5	98.9	631	9 US-09-795-926-41	Sequence 41, Appl
5	3242.5	98.9	631	14 US-10-364-774-41	Sequence 41, Appl
6	2771	84.5	506	9 US-09-795-926-31	Sequence 31, Appl
7	2771	84.5	506	14 US-10-364-774-31	Sequence 31, Appl
8	2746.5	83.8	535	9 US-09-795-926-29	Sequence 29, Appl
9	2746.5	83.8	535	14 US-10-364-774-29	Sequence 29, Appl
10	1877.	57.3	366	9 US-09-795-926-39	Sequence 39, Appl
11	1877.	57.3	366	14 US-10-364-774-39	Sequence 39, Appl
12	1392	42.5	276	14 US-10-292-836-66	Sequence 3, Appl
13	1370	41.8	269	9 US-09-795-926-27	Sequence 27, Appl



QY 241 RANVNLWPLLDIARNKTIIVCPMIDVHDIDRRETOAGDAMRGAFDEMWTYKRCPIP 300  
 Db 1 MRRKEKRILQAVAVLAVLVLNVGLWALYRERQPGCTPGGSGAAVAPAGCOSHRSQK 60  
 Db 241 EANVNLWPLLDIARNKTIIVCPMIDVHDIDRRETOAGDAMRGAFDEMWTYKRCPIP 300  
 QY 301 PELOKADPSDPFESPVAGLFAVDRKWFELGGYDGLBINGESEQYTSFKWMCGRM 360  
 Db 121 ISNRSIDPDRHNCNSKYLETLPLNTSIIIPHNEGQSSLLRTVHSYLNRSPELVEI 180  
 QY 361 EDIPCSRQHIIYRKYKVPYKVPAGVSLARNLKRVAEVNMDAYEAYIYORRPEYHLSAGD 420  
 Db 121 ISNRSIDPDRHNCNSKYLETLPLNTSIIIPHNEGQSSLLRTVHSYLNRSPELVEI 180  
 Db 361 EDIPCSRQHIIYRKYKVPYKVPAGVSLARNLKRVAEVNMDAYEAYIYORRPEYHLSAGD 420  
 QY 421 AVQKURSSINCKSPKWMKIAWDLPKTPYKVPYKVPAGVSLARNLKRVAEVNMDAYEAYIYORRPEYHLSAGD 480  
 Db 181 VLVDPSPREHLLKPLEDMALFSPVRLTRKREGRLTRMAGASVATGDVITFLDSHC 240  
 QY 421 AVOKURSSINCKSPKWMKIAWDLPKTPYKVPYKVPAGVSLARNLKRVAEVNMDAYEAYIYORRPEYHLSAGD 480  
 Db 181 VLVDPSPREHLLKPLEDMALFSPVRLTRKREGRLTRMAGASVATGDVITFLDSHC 240  
 Db 481 SPARLEGCVGRGRGAAWMNMQVFTWREDIRGDPOHITKCFDAISHTSPVLYDHS 540  
 QY 541 MKGNOLWQYKDKYLYHPSGSCMDCSDHRIFMNTCPSSUQOMFEHTNSTVLPK 600  
 Db 301 PELOKADPSDPFESPVAGLFAVDRKWFELGGYDGLBINGESEQYTSFKWMCGRM 360  
 QY 541 MKGNOLWQYKDKYLYHPSGSCMDCSDHRIFMNTCPSSUQOMFEHTNSTVLPK 600  
 Db 361 EDIPCSRQHIIYRKYKVPYKVPAGVSLARNLKRVAEVNMDAYEAYIYORRPEYHLSAGD 420  
 QY 601 NRN 603  
 Db 601 NRN 603

RESULT 3

US-10-364-774-43  
 ; Sequence 43, Application US/10364774  
 ; Publication No. US20030144497A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Donoho, Gregory  
 ; APPLICANT: Turner, C. Alexander Jr.  
 ; APPLICANT: Friedrich, Glenn  
 ; APPLICANT: Abuin, Alejandro  
 ; APPLICANT: Zambowicz, Brian  
 ; APPLICANT: Sands, Arthur T.  
 ; APPLICANT: Wilgowski, Nathaniel L.  
 ; APPLICANT: Hu, Yi  
 ; APPLICANT: Kleke, James Alvin  
 ; APPLICANT: Potter, David George  
 ; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
 ; FILE REFERENCE: LEX-0144-USA  
 ; CURRENT APPLICATION NUMBER: US/10/364,774  
 ; CURRENT FILING DATE: 2003-02-11  
 ; PRIOR APPLICATION NUMBER: US/09/795,926  
 ; PRIOR FILING DATE: 2001-02-28  
 ; PRIOR APPLICATION NUMBER: US/09/795,920  
 ; PRIOR FILING DATE: 2000-02-29  
 ; PRIOR APPLICATION NUMBER: US/60/186,558  
 ; PRIOR FILING DATE: 2000-03-02  
 ; PRIOR APPLICATION NUMBER: US/60/191,849  
 ; PRIOR FILING DATE: 2000-03-24  
 ; NUMBER OF SEQ ID NOS: 47  
 ; SOFTWARE: FastaSEQ for Windows Version 4.0  
 ; SEQ ID NO 43  
 ; LENGTH: 603  
 ; TYPE: PRT  
 ; ORGANISM: homo sapiens

QY 481 SPARLEGCVGRGRGAAWMNMQVFTWREDIRGDPOHITKCFDAISHTSPVLYDHS 540  
 Db 481 SPARLEGCVGRGRGAAWMNMQVFTWREDIRGDPOHITKCFDAISHTSPVLYDHS 540  
 QY 541 MKGNOLWQYKDKYLYHPSGSCMDCSDHRIFMNTCPSSUQOMFEHTNSTVLPK 600  
 Db 301 PELOKADPSDPFESPVAGLFAVDRKWFELGGYDGLBINGESEQYTSFKWMCGRM 360  
 QY 541 MKGNOLWQYKDKYLYHPSGSCMDCSDHRIFMNTCPSSUQOMFEHTNSTVLPK 600  
 Db 361 EDIPCSRQHIIYRKYKVPYKVPAGVSLARNLKRVAEVNMDAYEAYIYORRPEYHLSAGD 420  
 QY 421 AVOKURSSINCKSPKWMKIAWDLPKTPYKVPYKVPAGVSLARNLKRVAEVNMDAYEAYIYORRPEYHLSAGD 480  
 Db 421 AVOKURSSINCKSPKWMKIAWDLPKTPYKVPYKVPAGVSLARNLKRVAEVNMDAYEAYIYORRPEYHLSAGD 480  
 QY 481 SPARLEGCVGRGRGAAWMNMQVFTWREDIRGDPOHITKCFDAISHTSPVLYDHS 540  
 Db 481 SPARLEGCVGRGRGAAWMNMQVFTWREDIRGDPOHITKCFDAISHTSPVLYDHS 540  
 QY 541 MKGNOLWQYKDKYLYHPSGSCMDCSDHRIFMNTCPSSUQOMFEHTNSTVLPK 600  
 Db 301 PELOKADPSDPFESPVAGLFAVDRKWFELGGYDGLBINGESEQYTSFKWMCGRM 360  
 QY 541 MKGNOLWQYKDKYLYHPSGSCMDCSDHRIFMNTCPSSUQOMFEHTNSTVLPK 600  
 Db 361 EDIPCSRQHIIYRKYKVPYKVPAGVSLARNLKRVAEVNMDAYEAYIYORRPEYHLSAGD 420  
 QY 601 NRN 603  
 Db 601 NRN 603

RESULT 4

US-09-795-926-41  
 ; Sequence 41, Application US/09795926  
 ; Patent No. US20020058486A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Donoho, Gregory  
 ; APPLICANT: Hilbun, Erin  
 ; APPLICANT: Turner, C. Alexander Jr.  
 ; APPLICANT: Friedrich, Glenn  
 ; APPLICANT: Abuin, Alejandro  
 ; APPLICANT: Zambowicz, Brian  
 ; APPLICANT: Sands, Arthur T.  
 ; APPLICANT: Wilgowski, Nathaniel L.  
 ; APPLICANT: Hu, Yi  
 ; APPLICANT: Kleke, James Alvin  
 ; APPLICANT: Potter, David George  
 ; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
 ; FILE REFERENCE: LEX-0144-USA  
 ; CURRENT APPLICATION NUMBER: US/09/795,926  
 ; CURRENT FILING DATE: 2001-02-28  
 ; PRIOR APPLICATION NUMBER: US/09/795,920  
 ; PRIOR FILING DATE: 2000-02-29  
 ; PRIOR APPLICATION NUMBER: US/60/186,558  
 ; PRIOR FILING DATE: 2000-03-02  
 ; PRIOR APPLICATION NUMBER: US/60/191,849  
 ; PRIOR FILING DATE: 2000-03-24  
 ; NUMBER OF SEQ ID NOS: 47  
 ; SOFTWARE: FastaSEQ for Windows Version 4.0  
 ; SEQ ID NO 43  
 ; LENGTH: 603  
 ; TYPE: PRT  
 ; ORGANISM: homo sapiens

Query Match 100.0%; Score 3278; DB 14; Length 603;  
 Best Local Similarity 100.0%; Pred. No. 5e-310; Mismatches 0; Indels 0; Gaps 0;  
 Matches 603; Conservative 0; MisMatches 0; Indels 0; Gaps 0;

QY 1 MRRKEKRILQAVAVLAVLVLPLWGLWLYRERQPGCTPGGSGAAVAPAGCOSHRSQK 60

```

; SEQ ID NO: 41
; LENGTH: 631
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-41

Query Match 98.9%; Score 3242.5; DB 9; Length 631;
Best Local Similarity 95.4%; Pred. No. 1.5e-306; Mismatches 0; Indels 29; Gaps 1;
Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

Qy 1 MRRKEKRLLQAVALVALVLPNGLWALYREROPDGPGGSGAAVAPAGQSHSRQK 60
Db 1 MRRKEKRLLQAVALVALVLPNGLWALYREROPDGPGGSGAAVAPAGQSHSRQK 60
Qy 61 KTFPLGDGQKLUKWDHKDEAIRRDAQRVGNGEQGRPYPMIDAEVRQDQAYRENGFNIYVSDK 120
Db 61 KTFPLGDGQKLUKWDHKDEAIRRDAQRVGNGEQGRPYPMIDAEVRQDQAYRENGFNIYVSDK 120
Qy 121 ISLRSRSLPDRPHPNCSKRYLETLNTSIIIPFHNEGWSLLRTVHSVUNRSPPELVET 180
Db 121 ISLRSRSLPDRPHPNCSKRYLETLNTSIIIPFHNEGWSLLRTVHSVUNRSPPELVET 180
Qy 181 VLVDDPSDREHLKKEPLEDMALFSPVRLTTRKREGRLRTRMLGASVATGDVITFLDSHC 240
Db 181 VLVDDPSDREHLKKEPLEDMALFSPVRLTTRKREGRLRTRMLGASVATGDVITFLDSHC 240
Qy 241 EANVNWLPPLDRIARNRKTIVCPMDVIDHDDFRYETOQAGDAMRGAFDWEEMYKRIPIP 300
Db 241 EANVNWLPPLDRIARNRKTIVCPMDVIDHDDFRYETOQAGDAMRGAFDWEEMYKRIPIP 300
Qy 301 PELQKADPSDPFESPVWAGLFLAVDRKWFELWLGYPDGLBWIWGEQEYLSFKGLMLPRL 360
Db 301 PELQKADPSDPFESPVWAGLFLAVDRKWFELWLGYPDGLBWIWGEQEYLSFKGLMLPRL 360
Qy 353 -----VWCGGRMEDIPCSRUGHYRKVYKPYKPGVAGVSLARNIK 391
Db 353 -----VWCGGRMEDIPCSRUGHYRKVYKPYKPGVAGVSLARNIK 391
Qy 361 VSNSWPOAVFLPRAFMALQWVNGGGRMEDIPCSRUGHYRKVYKPYKPGVAGVSLARNIK 420
Db 361 VSNSWPOAVFLPRAFMALQWVNGGGRMEDIPCSRUGHYRKVYKPYKPGVAGVSLARNIK 420
Qy 392 RVAEWMDEVAEYIYQRRPRYRHLASAGDVAQVKURSSLNCKSKPKWFMKTLAIDLPKFYP 451
Db 421 RVAEWMDEVAEYIYQRRPRYRHLASAGDVAQVKURSSLNCKSKPKWFMKTLAIDLPKFYP 480
Qy 452 PVEPPAANGEIIRVNGTGLCADTRKGALGSPLRFLRGCVRGGRGEAANNNNQVFTWREDI 511
Db 481 PVEPPAANGEIIRVNGTGLCADTRKGALGSPLRFLRGCVRGGRGEAANNNNQVFTWREDI 540
Qy 512 RPGDQHTKRCFCDAISHTSPVTLYDCHSMKGNGQLWKYRKDKTLIYHPVGSCMCSESDH 571
Db 541 RPGDQHTKRCFCDAISHTSPVTLYDCHSMKGNGQLWKYRKDKTLIYHPVGSCMCSESDH 600
Qy 572 RIFMTCNFSSLTQWLFHTNSTLEKEN 601
Db 601 RIFMTCNFSSLTQWLFHTNSTLEKEN 630

RESULT 5
US-10-364-774-41
; Sequence 41, Application US/10364774
; Publication No. US20030144497A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Grin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sande, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilgnowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
; FILE REFERENCE: DEK-0144-USA
; CURRENT APPLICATION: US/10/364,774
; CURRENT FILING DATE: 2003-02-11
; PRIORITY NUMBER: US/09/795,926
; PRIORITY FILING DATE: 2001-02-28
; PRIORITY APPLICATION NUMBER: US 60/185,920
; PRIORITY FILING DATE: 2000-02-29
; PRIORITY APPLICATION NUMBER: US 60/186,558
; PRIORITY APPLICATION NUMBER: US 60/191,849
; PRIORITY FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO: 41
; LENGTH: 631
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-364-774-41

Query Match 98.9%; Score 3242.5; DB 14; Length 631;
Best Local Similarity 95.4%; Pred. No. 1.5e-305; Mismatches 0; Indels 29; Gaps 1;
Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

Qy 1 MRRKEKRLLQAVALVALVLPNGLWALYREROPDGPGGSGAAVAPAGQSHSRQK 60
Db 1 MRRKEKRLLQAVALVALVLPNGLWALYREROPDGPGGSGAAVAPAGQSHSRQK 60
Qy 61 KTFPLGDGQKLUKWDHKDEAIRRDAQRVGNGEQGRPYPMIDAEVRQDQAYRENGFNIYVSDK 120
Db 61 KTFPLGDGQKLUKWDHKDEAIRRDAQRVGNGEQGRPYPMIDAEVRQDQAYRENGFNIYVSDK 120
Qy 121 ISLRSRSLPDRPHPNCSKRYLETLNTSIIIPFHNEGWSLLRTVHSVUNRSPPELVET 180
Db 121 ISLRSRSLPDRPHPNCSKRYLETLNTSIIIPFHNEGWSLLRTVHSVUNRSPPELVET 180
Qy 181 VLVDDPSDREHLKKEPLEDMALFSPVRLTTRKREGRLRTRMLGASVATGDVITFLDSHC 240
Db 181 VLVDDPSDREHLKKEPLEDMALFSPVRLTTRKREGRLRTRMLGASVATGDVITFLDSHC 240
Qy 241 EANVNWLPPLDRIARNRKTIVCPMDVIDHDDFRYETOQAGDAMRGAFDWEEMYKRIPIP 300
Db 241 EANVNWLPPLDRIARNRKTIVCPMDVIDHDDFRYETOQAGDAMRGAFDWEEMYKRIPIP 300
Qy 301 PELQKADPSDPFESPVWAGLFLAVDRKWFELWLGYPDGLBWIWGEQEYLSFKGLMLPRL 360
Db 301 PELQKADPSDPFESPVWAGLFLAVDRKWFELWLGYPDGLBWIWGEQEYLSFKGLMLPRL 360
Qy 353 -----VWCGGRMEDIPCSRUGHYRKVYKPYKPGVAGVSLARNIK 391
Db 353 -----VWCGGRMEDIPCSRUGHYRKVYKPYKPGVAGVSLARNIK 391
Qy 361 VSNSWPOAVFLPRAFMALQWVNGGGRMEDIPCSRUGHYRKVYKPYKPGVAGVSLARNIK 420
Db 361 VSNSWPOAVFLPRAFMALQWVNGGGRMEDIPCSRUGHYRKVYKPYKPGVAGVSLARNIK 420
Qy 392 RVAEWMDEVAEYIYQRRPRYRHLASAGDVAQVKURSSLNCKSKPKWFMKTLAIDLPKFYP 451
Db 421 RVAEWMDEVAEYIYQRRPRYRHLASAGDVAQVKURSSLNCKSKPKWFMKTLAIDLPKFYP 480
Qy 452 PVEPPAANGEIIRVNGTGLCADTRKGALGSPLRFLRGCVRGGRGEAANNNNQVFTWREDI 511
Db 481 PVEPPAANGEIIRVNGTGLCADTRKGALGSPLRFLRGCVRGGRGEAANNNNQVFTWREDI 540
Qy 512 RPGDQHTKRCFCDAISHTSPVTLYDCHSMKGNGQLWKYRKDKTLIYHPVGSCMCSESDH 571
Db 541 RPGDQHTKRCFCDAISHTSPVTLYDCHSMKGNGQLWKYRKDKTLIYHPVGSCMCSESDH 600
Qy 572 RIFMTCNFSSLTQWLFHTNSTLEKEN 601
Db 601 RIFMTCNFSSLTQWLFHTNSTLEKEN 630

RESULT 6
US-09-795-926-31
; Sequence 31, Application US/09795926
; Patent No. US20020098486A1

```

; GENERAL INFORMATION:  
; APPLICANT: Denoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilgowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Klek, James Alvin  
; APPLICANT: Porter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/09/795,926  
; CURRENT FILING DATE: 2000-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 31  
; LENGTH: 506  
; TYPE: PRT  
; ORGANISM: homo sapiens  
; US-09-795-926-31

Query Match 84.5%; Score 2771; DB 9; Length 506;  
Best Local Similarity 100.0%; Pred. No. 1e-260; Mismatches 0; Indels 0; Gaps 0;  
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 98 MTDAAERVQDQAYRENGFNYVSDKISLNLSPDTRHPNCNSKRYLETNTSIIIPFNEG 157  
1 MTDAAERVQDQAYRENGFNYVSDKISLNLSPDTRHPNCNSKRYLETNTSIIIPFNEG 60

Db 158 WSSLRRTVHSVLRSPPELVAELVLDPSDRHLKKELEDWALFSPVRILTKKREGL 217  
61 WSSLRRTVHSVLRSPPELVAELVLDPSDRHLKKELEDWALFSPVRILTKKREGL 120

Qy 218 IRTTMLGASVATGDPVITLDSCHEANTWLPILLDRARNRKTIVCPMIDVHDFFRYE 277  
121 IRTTMLGASVATGDPVITLDSCHEANTWLPILLDRARNRKTIVCPMIDVHDFFRYE 180

Qy 278 TOQDAMRGAFDNEMYKRIPIPPELOKADPSPDFESPVMAGGLFAVDRKWWELGGDP 337  
181 TOQDAMRGAFDNEMYKRIPIPPELOKADPSPDFESPVMAGGLFAVDRKWWELGGDP 240

Db 338 GLIJIWGGQYEFSPKVNMGGRMEDICPSRVSRYHRYKVKPYKPGAVSLARNLKVRAEW 397  
241 GLIJIWGGQYEFSPKVNMGGRMEDICPSRVSRYHRYKVKPYKPGAVSLARNLKVRAEW 300

Qy 398 MDYAEYIYQRRPEYRHLSSAGDVAVOKLRSNLNSCKSFWMKIAMLPLKYPVPPERA 457  
301 MDYAEYIYQRRPEYRHLSSAGDVAVOKLRSNLNSCKSFWMKIAMLPLKYPVPPERA 360

Db 458 AANGBEIRNGTGLCADTKGAGLGSPLRLEGTCVGRGERGAEMWNMQVFTTWRDIDRPGDQ 517  
361 AANGBEIRNGTGLCADTKGAGLGSPLRLEGTCVGRGERGAEMWNMQVFTTWRDIDRPGDQ 420

Db 518 HTKKFCFDAISHTSPVTLYDCHSMKGNLWKYRKDKTLYHPSGSCMDCSSESDHRIFTNT 577  
421 HTKKFCFDAISHTSPVTLYDCHSMKGNLWKYRKDKTLYHPSGSCMDCSSESDHRIFTNT 480

Qy 578 CNSSLTQOWLFRNTSTLKEFTRN 603  
481 CNSSLTQOWLFRNTSTLKEFTRN 506

RESULT 7  
US-10-364-774-31  
Sequence 31, Application US/10364774  
Publication No. US20030144497A1  
; GENERAL INFORMATION:  
; APPLICANT: Denoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilgowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Klek, James Alvin  
; APPLICANT: Porter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/10/364,774  
; CURRENT FILING DATE: 2003-03-11  
; PRIOR APPLICATION NUMBER: US/09/795,926  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 31  
; LENGTH: 506  
; TYPE: PRT  
; ORGANISM: homo sapiens  
; US-10-364-774-31

Query Match 84.5%; Score 2771; DB 14; Length 506;  
Best Local Similarity 100.0%; Pred. No. 1e-260; Mismatches 0; Indels 0; Gaps 0;  
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 98 MTDAAERVQDQAYRENGFNYVSDKISLNLSPDTRHPNCNSKRYLETNTSIIIPFNEG 157  
1 MTDAAERVQDQAYRENGFNYVSDKISLNLSPDTRHPNCNSKRYLETNTSIIIPFNEG 60

Db 158 WSSLRRTVHSVLRSPPELVAELVLDPSDRHLKKELEDWALFSPVRILTKKREGL 217  
61 WSSLRRTVHSVLRSPPELVAELVLDPSDRHLKKELEDWALFSPVRILTKKREGL 120

Qy 218 IRTTMLGASVATGDPVITLDSCHEANTWLPILLDRARNRKTIVCPMIDVHDFFRYE 277  
121 IRTTMLGASVATGDPVITLDSCHEANTWLPILLDRARNRKTIVCPMIDVHDFFRYE 180

Qy 278 TOQDAMRGAFDNEMYKRIPIPPELOKADPSPDFESPVMAGGLFAVDRKWWELGGDP 337  
181 TOQDAMRGAFDNEMYKRIPIPPELOKADPSPDFESPVMAGGLFAVDRKWWELGGDP 240

Db 338 GLIJIWGGQYEFSPKVNMGGRMEDICPSRVSRYHRYKVKPYKPGAVSLARNLKVRAEW 397  
241 GLIJIWGGQYEFSPKVNMGGRMEDICPSRVSRYHRYKVKPYKPGAVSLARNLKVRAEW 300

Qy 398 MDYAEYIYQRRPEYRHLSSAGDVAVOKLRSNLNSCKSFWMKIAMLPLKYPVPPERA 457  
301 MDYAEYIYQRRPEYRHLSSAGDVAVOKLRSNLNSCKSFWMKIAMLPLKYPVPPERA 360

Db 458 AANGBEIRNGTGLCADTKGAGLGSPLRLEGTCVGRGERGAEMWNMQVFTTWRDIDRPGDQ 517  
361 AANGBEIRNGTGLCADTKGAGLGSPLRLEGTCVGRGERGAEMWNMQVFTTWRDIDRPGDQ 420

Db 518 HTKKFCFDAISHTSPVTLYDCHSMKGNLWKYRKDKTLYHPSGSCMDCSSESDHRIFTNT 577  
421 HTKKFCFDAISHTSPVTLYDCHSMKGNLWKYRKDKTLYHPSGSCMDCSSESDHRIFTNT 480

Qy 398 MDYAEYIYQRRPEYRHLSSAGDVAVOKLRSNLNSCKSFWMKIAMLPLKYPVPPERA 457  
301 MDYAEYIYQRRPEYRHLSSAGDVAVOKLRSNLNSCKSFWMKIAMLPLKYPVPPERA 360

Db 458 AANGBEIRNGTGLCADTKGAGLGSPLRLEGTCVGRGERGAEMWNMQVFTTWRDIDRPGDQ 517  
361 AANGBEIRNGTGLCADTKGAGLGSPLRLEGTCVGRGERGAEMWNMQVFTTWRDIDRPGDQ 420

Qy 518 HTKKFCFDAISHTSPVTLYDCHSMKGNLWKYRKDKTLYHPSGSCMDCSSESDHRIFTNT 577  
421 HTKKFCFDAISHTSPVTLYDCHSMKGNLWKYRKDKTLYHPSGSCMDCSSESDHRIFTNT 480

RESULT 8

US-09-795-926-29

; Sequence 29, Application US/09795926

; Patent No. US2002008486A1

GENERAL INFORMATION:

APPLICANT: Donoho, Gregory

APPLICANT: Hilbun, Erin

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Friedrich, Glenn

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sands, Arthur T.

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Hu, Yi

APPLICANT: Kleke, James Alvin

APPLICANT: Potter, David George

APPLICANT: Potter, James Alvin

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sands, Arthur T.

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795,926

CURRENT FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: US 60/185, 920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186, 558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191, 849

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: Fast-SEQ for Windows Version 4.0

SEQ ID NO: 29

LENGTH: 535

TYPE: PRT

ORGANISM: homo sapiens

US-09-795-926-29

Query Match 83.8%; Score 2746.5; DB 9; Length 535; Best Local Similarity 94.6%; Pred. No. 2.8e-258; Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

Qy 98 WTDRAERVDQAYRENGFTYVSDKLSLNLSPDTRPHNCNSKRYLETPNTSTIIPFNEG 157

Db 1 MTDRAERVDQAYRENGFTYVSDKLSLNLSPDTRPHNCNSKRYLETPNTSTIIPFNEG 60

Qy 158 WSSLRRTVHSVLRNSPPLVAETVLUVDPDSREHLKKPLEDMLAFLPSVRIRTKRREGL 217

Db 61 WSSLRRTVHSVLRNSPPLVAETVLUVDPDSREHLKKPLEDMLAFLPSVRIRTKRREGL 120

Qy 218 ITRMLGASVATGDTVFLDSCEANVNLPLLDRAARNKRTIVCPMDV1DHDPFRE 277

Db 121 ITRMLGASVATGDTVFLDSCEANVNLPLLDRAARNKRTIVCPMDV1DHDPFRE 180

Qy 278 TQGDAERGAFDPEMYYKRIPPELQKADPSDPFPSPVMAGGLPAVDRKWFELGGYDP 337

Db 181 TQGDAERGAFDPEMYYKRIPPELQKADPSDPFPSPVMAGGLPAVDRKWFELGGYDP 240

Qy 338 GLEIWWGQEOYESPK-----VWMCGRMEDICPSRV 368

Db 241 GLEIWWGQEOYESPKGLMLPRLVSNWPOAVFLPRAPNMLQVMCGGRMEDICPSRV 300

Qy 369 GHITRKVYKPVAGVSLARNLKRVAEWMDDEAYTIVQRREYRHLISAGDVAVQKLLRS 428

Db 301 GHITRKVYKPVAGVSLARNLKRVAEWMDDEAYTIVQRREYRHLISAGDVAVQKLLRS 360

Qy 429 SLMCKSFKNMFTKLAIDLKPKFYPPVPPAAGKSEIRNVTGICADTKIGALSPRLLEG 488

Db 361 SLMCKSFKNMFTKLAIDLKPKFYPPVPPAAGKSEIRNVTGICADTKIGALSPRLLEG 420

RESULT 9

US-10-364-774-29

; Sequence 29, Application US/10364774

Publication No. US2003014447A1

GENERAL INFORMATION:

APPLICANT: Donoho, Gregory

APPLICANT: Hilbun, Erin

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Friedrich, Glenn

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sands, Arthur T.

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Hu, Yi

APPLICANT: Kleke, James Alvin

APPLICANT: Potter, David George

APPLICANT: Potter, James Alvin

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sands, Arthur T.

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795,926

CURRENT FILING DATE: 2003-02-11

PRIOR APPLICATION NUMBER: US/09/795,926

PRIOR FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: US 60/185, 920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186, 558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191, 849

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: Fast-SEQ for Windows Version 4.0

SEQ ID NO: 29

LENGTH: 535

TYPE: PRT

ORGANISM: homo sapiens

US-10-364-774-29

Query Match 83.8%; Score 2746.5; DB 14; Length 535; Best Local Similarity 94.6%; Pred. No. 2.8e-258; Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

Qy 98 WTDRAERVDQAYRENGFTYVSDKLSLNLSPDTRPHNCNSKRYLETPNTSTIIPFNEG 157

Db 1 MTDRAERVDQAYRENGFTYVSDKLSLNLSPDTRPHNCNSKRYLETPNTSTIIPFNEG 60

Qy 158 WSSLRRTVHSVLRNSPPLVAETVLUVDPDSREHLKKPLEDMLAFLPSVRIRTKRREGL 217

Db 61 WSSLRRTVHSVLRNSPPLVAETVLUVDPDSREHLKKPLEDMLAFLPSVRIRTKRREGL 120

Qy 218 ITRMLGASVATGDTVFLDSCEANVNLPLLDRAARNKRTIVCPMDV1DHDPFRE 277

Db 121 ITRMLGASVATGDTVFLDSCEANVNLPLLDRAARNKRTIVCPMDV1DHDPFRE 180

Qy 278 TQGDAERGAFDPEMYYKRIPPELQKADPSDPFPSPVMAGGLPAVDRKWFELGGYDP 337

Db 181 TQGDAERGAFDPEMYYKRIPPELQKADPSDPFPSPVMAGGLPAVDRKWFELGGYDP 240

Qy 338 GLEIWWGQEOYESPK-----VWMCGRMEDICPSRV 368

Db 241 GLEIWWGQEOYESPKGLMLPRLVSNWPOAVFLPRAPNMLQVMCGGRMEDICPSRV 300

Qy 369 GHITRKVYKPVAGVSLARNLKRVAEWMDDEAYTIVQRREYRHLISAGDVAVQKLLRS 428

Db 301 GHITRKVYKPVAGVSLARNLKRVAEWMDDEAYTIVQRREYRHLISAGDVAVQKLLRS 360

Qy 429 SLMCKSFKNMFTKLAIDLKPKFYPPVPPAAGKSEIRNVTGICADTKIGALSPRLLEG 488

Db 361 SLMCKSFKNMFTKLAIDLKPKFYPPVPPAAGKSEIRNVTGICADTKIGALSPRLLEG 420

RESULT 10

US-10-364-774-29

; Sequence 29, Application US/10364774

Publication No. US2003014447A1

GENERAL INFORMATION:

APPLICANT: Donoho, Gregory

APPLICANT: Hilbun, Erin

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Friedrich, Glenn

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sands, Arthur T.

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Hu, Yi

APPLICANT: Kleke, James Alvin

APPLICANT: Potter, David George

APPLICANT: Potter, James Alvin

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sands, Arthur T.

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795,926

CURRENT FILING DATE: 2003-02-11

PRIOR APPLICATION NUMBER: US/09/795,926

PRIOR FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: US 60/185, 920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186, 558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191, 849

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: Fast-SEQ for Windows Version 4.0

SEQ ID NO: 29

LENGTH: 535

TYPE: PRT

ORGANISM: homo sapiens

US-10-364-774-29

Query Match 83.8%; Score 2746.5; DB 14; Length 535; Best Local Similarity 94.6%; Pred. No. 2.8e-258; Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

Qy 98 WTDRAERVDQAYRENGFTYVSDKLSLNLSPDTRPHNCNSKRYLETPNTSTIIPFNEG 157

Db 1 MTDRAERVDQAYRENGFTYVSDKLSLNLSPDTRPHNCNSKRYLETPNTSTIIPFNEG 60

Qy 158 WSSLRRTVHSVLRNSPPLVAETVLUVDPDSREHLKKPLEDMLAFLPSVRIRTKRREGL 217

Db 61 WSSLRRTVHSVLRNSPPLVAETVLUVDPDSREHLKKPLEDMLAFLPSVRIRTKRREGL 120

Qy 218 ITRMLGASVATGDTVFLDSCEANVNLPLLDRAARNKRTIVCPMDV1DHDPFRE 277

Db 121 ITRMLGASVATGDTVFLDSCEANVNLPLLDRAARNKRTIVCPMDV1DHDPFRE 180

Qy 278 TQGDAERGAFDPEMYYKRIPPELQKADPSDPFPSPVMAGGLPAVDRKWFELGGYDP 337

Db 181 TQGDAERGAFDPEMYYKRIPPELQKADPSDPFPSPVMAGGLPAVDRKWFELGGYDP 240

Qy 338 GLEIWWGQEOYESPK-----VWMCGRMEDICPSRV 368

Db 241 GLEIWWGQEOYESPKGLMLPRLVSNWPOAVFLPRAPNMLQVMCGGRMEDICPSRV 300

Qy 369 GHITRKVYKPVAGVSLARNLKRVAEWMDDEAYTIVQRREYRHLISAGDVAVQKLLRS 428

RESULT 10  
US-09-795-926-39  
; Sequence 39, Application US/09/795,926  
; Patent No. US20020058486A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrovic, Brian  
; APPLICANT: Sands, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilgnowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kleke, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/09/795,926  
; CURRENT FILING DATE: 2001-02-28  
; CURRENT APPLICATION NUMBER: US/09/795,926  
; CURRENT FILING DATE: 2001-02-28  
; PRIORITY NUMBER: US 60/185,920  
; PRIORITY NUMBER: US 60/186,558  
; PRIORITY FILING DATE: 2000-02-29  
; PRIORITY FILING DATE: 2000-03-02  
; PRIORITY APPLICATION NUMBER: US 60/191,849  
; PRIORITY FILING DATE: 2000-03-24  
; PRIORITY NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 39  
; LENGTH: 366  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-795-926-39  
; Query Match 57.3%; Score 1877; DB 9; Length 366;  
; Best Local Similarity 100.0%; Pred. No. 9.5e-14;  
; Mismatches 0; Indels 0; Gaps 0;  
; Matches 353; Conservative 0;  
; QY 1 MRKEKRLQAVVALVLAALVLPLNGJMLWYRQDQPGSGAAVAPAGQSHSRQK 60  
; DB 1 MRKEKRLQAVVALVLAALVLPLNGJMLWYRQDQPGSGAAVAPAGQSHSRQK 60  
; QY 61 KTFPLGDKQKWKWDKEAIRRDAQRVNGEQQGPPWMDAERVDQAVRENGNIVYSDK 120  
; DB 61 KTFPLGDKQKWKWDKEAIRRDAQRVNGEQQGPPWMDAERVDQAVRENGNIVYSDK 120  
; QY 121 ISLNRSLPDIRHNCNSKRYLETPLNTSIIIPFHNEGMSLRTVHVSINRSPPELVAEI 180  
; DB 121 ISLNRSLPDIRHNCNSKRYLETPLNTSIIIPFHNEGMSLRTVHVSINRSPPELVAEI 180  
; QY 181 VLVDDPSREHLKPLEDYMALPSVRLRTKREGRLTRMGASVATGDTVFLSHC 240  
; DB 181 VLVDDPSREHLKPLEDYMALPSVRLRTKREGRLTRMGASVATGDTVFLSHC 240  
; QY 241 BANVWLPPLLDRIARNRKTIVCPMIDVIDHDDPRTYETOAGDAMGAFDWEMYKRIPIP 300  
; DB 241 BANVWLPPLLDRIARNRKTIVCPMIDVIDHDDPRTYETOAGDAMGAFDWEMYKRIPIP 300  
; QY 301 PELOKADSDPFPSPVMGGFLAVDRKMFWEGLGGYDGEGLTGWGEYEFISFKV 353  
; DB 301 PELOKADSDPFPSPVMGGFLAVDRKMFWEGLGGYDGEGLTGWGEYEFISFKV 353

RESULT 11  
US-10-364-774-39  
; Sequence 39, Application US/10/364,774  
; Publication No. US20030144971A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrovic, Brian  
; APPLICANT: Sands, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilgnowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kleke, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/10/364,774  
; CURRENT FILING DATE: 2003-02-11  
; PRIORITY NUMBER: US/09/795,926  
; PRIORITY NUMBER: US 60/185,920  
; PRIORITY NUMBER: US 60/186,558  
; PRIORITY FILING DATE: 2000-03-02  
; PRIORITY APPLICATION NUMBER: US 60/191,849  
; PRIORITY FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 39  
; LENGTH: 366  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-364-774-39  
; Query Match 57.3%; Score 1877; DB 14; Length 366;  
; Best Local Similarity 100.0%; Pred. No. 9.5e-14;  
; Mismatches 0; Indels 0; Gaps 0;  
; Matches 353; Conservative 0;  
; QY 1 MRKEKRLQAVVALVLAALVLPLNGJMLWYRQDQPGSGAAVAPAGQSHSRQK 60  
; DB 1 MRKEKRLQAVVALVLAALVLPLNGJMLWYRQDQPGSGAAVAPAGQSHSRQK 60  
; QY 61 KTFPLGDKQKWKWDKEAIRRDAQRVNGEQQGPPWMDAERVDQAVRENGNIVYSDK 120  
; DB 61 KTFPLGDKQKWKWDKEAIRRDAQRVNGEQQGPPWMDAERVDQAVRENGNIVYSDK 120  
; QY 121 ISLNRSLPDIRHNCNSKRYLETPLNTSIIIPFHNEGMSLRTVHVSINRSPPELVAEI 180  
; DB 121 ISLNRSLPDIRHNCNSKRYLETPLNTSIIIPFHNEGMSLRTVHVSINRSPPELVAEI 180  
; QY 181 VLVDDPSREHLKPLEDYMALPSVRLRTKREGRLTRMGASVATGDTVFLSHC 240  
; DB 181 VLVDDPSREHLKPLEDYMALPSVRLRTKREGRLTRMGASVATGDTVFLSHC 240  
; QY 301 PELOKADSDPFPSPVMGGFLAVDRKMFWEGLGGYDGEGLTGWGEYEFISFKV 353  
; DB 301 PELOKADSDPFPSPVMGGFLAVDRKMFWEGLGGYDGEGLTGWGEYEFISFKV 353

RESULT 12  
 US-10-292-896-3  
 Sequence 3, Application US/10292896  
 Publication No. US20030186650A1  
 GENERAL INFORMATION:  
 APPLICANT: HASSAN, Helle  
 APPLICANT: REIS, Celso A.  
 APPLICANT: BENNETT, Eric P.  
 APPLICANT: CLAUSEN, Henrik  
 TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GR  
 TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS  
 TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS  
 FILE REFERENCE: 4305/1H14-US3  
 CURRENT APPLICATION NUMBER: US/10/292,896  
 CURRENT FILING DATE: 2003-03-17  
 CURRENT FILING DATE: 2003-03-17  
 PRIOR APPLICATION NUMBER: US 60/186,504  
 PRIOR FILING DATE: 2002-11-08  
 PRIOR APPLICATION NUMBER: PCT/DK01/00328  
 PRIOR FILING DATE: 2001-03-10  
 PRIOR APPLICATION NUMBER: US 60/203,331  
 PRIOR FILING DATE: 2000-05-11  
 NUMBER OF SEQ ID NOS: 130  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO: 3  
 LENGTH: 276  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-292-896-3

Query Match 42.5%; Score 1392; DB 14; Length 276;  
 Best Local Similarity 97.3%; Pred. No. 1.2e-126;  
 Matches 250; Conservative 3; Mismatches 2; Gaps 1;  
 Qy 349 ISPKWMMGGAGMEDIPCSRQVGHIIKRVPKVVKPAGVSLA--RNLRVABEWMDAYKIV 406  
 Db 20 LAMQTMCGGAGMEDIPCSRQVGHIIKRVPKVVKPAGVSLA--RNLRVABEWMDAYKIV 79  
 Qy 407 ORRPFYRHTSAGDVAVOKLRSIINCKSKFWMFTKIAWDLKPKFIPVVERPAANGEBIRAV 466  
 Db 80 ORRPFYRHTSAGDVAVOKLRSIINCKSKFWMFTKIAWDLKPKFIPVVERPAANGEBIRAV 139  
 Qy 467 GTGLCADTKHGALASPLRLGCVRGERAANWNOVFTWREDIRGPQHTKFCFDA 526  
 Db 140 GTGLCADTKHGALASPLRLGCVRGERAANWNOVFTWREDIRGPQHTKFCFDA 199  
 Qy 527 ISHSHPVTLYDCHSMKGNOQWVKYRDKTLYHPVSGSCMDGSSDRHIFMTCNSSLTQQ 586  
 Db 200 ISHSHPVTLYDCHSMKGNOQWVKYRDKTLYHPVSGSCMDGSSDRHIFMTCNSSLTQQ 259  
 Qy 587 WLFEMHNSTVLEKENRN 603  
 Db 260 WLFEMHNSTVLEKENRN 276

RESULT 13  
 US-09-795-926-27  
 Sequence 27, Application US/09795926  
 Publication No. US2000009886A1  
 GENERAL INFORMATION:  
 APPLICANT: Donoho, Gregory  
 APPLICANT: Hilburn, Erin  
 APPLICANT: Turner, C. Alexander Jr.  
 APPLICANT: Friedrich, Glenn  
 APPLICANT: Abuin, Alejandro  
 APPLICANT: Zambrowicz, Brian  
 APPLICANT: Sands, Arthur T.  
 APPLICANT: Walke, D. Wade  
 APPLICANT: Wilgowski, Nathaniel L.  
 APPLICANT: Donoho, Gregory  
 APPLICANT: Hilburn, Erin  
 APPLICANT: Turner, C. Alexander Jr.  
 APPLICANT: Friedrich, Glenn  
 APPLICANT: Abuin, Alejandro  
 APPLICANT: Zambrowicz, Brian  
 APPLICANT: Sands, Arthur T.  
 APPLICANT: Walke, D. Wade  
 APPLICANT: Wilgowski, Nathaniel L.  
 APPLICANT: Kieke, James Alvin  
 APPLICANT: Potter, David George  
 TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
 TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
 FILE REFERENCE: LEX-0144-USA  
 CURRENT APPLICATION NUMBER: US/10/364,774  
 CURRENT FILING DATE: 2003-02-11  
 PRIOR APPLICATION NUMBER: US/09/795,926  
 PRIOR FILING DATE: 2001-02-28  
 PRIOR APPLICATION NUMBER: US 60/185,920  
 PRIOR FILING DATE: 2000-02-29  
 PRIOR APPLICATION NUMBER: US 60/186,553  
 PRIOR FILING DATE: 2000-03-02  
 PRIOR APPLICATION NUMBER: US 60/191,849  
 PRIOR FILING DATE: 2000-03-24  
 NUMBER OF SEQ ID NOS: 47  
 SEQ ID NO: 27  
 LENGTH: 269  
 TYPE: PRT  
 ORGANISM: homo sapiens  
 US-09-795-926-27

Query Match 41.8%; Score 1370; DB 9; Length 269;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-14; Matches 256; Conservative 0; Mismatches 0; Gaps 0;  
 Qy 98 MTDABRVDAYRENGFNIVVSDKISLNNSLSPDTRHPNNSKRYLETNTSIIIPFHNEG 157  
 Db 1 MTDABRVDAYRENGFNIVVSDKISLNNSLSPDTRHPNNSKRYLETNTSIIIPFHNEG 60  
 Qy 158 WSSLURTIVSVLNSRSPPIVAETVLDPSDRHMLKQLEDDYALPFSRVIARTKFCGL 217  
 Db 61 WSSLURTIVSVLNSRSPPIVAETVLDPSDRHMLKQLEDDYALPFSRVIARTKFCGL 120  
 Qy 218 IRTRMLGASVATGVITFLDSHEANVWLPPLDRINRNKTYCPCMIDVHDPRYE 277  
 Db 121 IRTRMLGASVATGVITFLDSHEANVWLPPLDRINRNKTYCPCMIDVHDPRYE 180  
 Qy 278 TQGADAMGAFDWMYKRIKIPPELQADPSPPFESVYMAGLFPAVDRKWFVWELGGDP 337  
 Db 181 TQGADAMGAFDWMYKRIKIPPELQADPSPPFESVYMAGLFPAVDRKWFVWELGGDP 240  
 Qy 338 GLEIWGGGEOYBISKY 353  
 Db 241 GLEIWGGGEOYBISKY 256

RESULT 14  
 US-10-364-774-27  
 Sequence 27, Application US/10364774  
 Publication No. US20030144497A1  
 GENERAL INFORMATION:  
 APPLICANT: Donoho, Gregory  
 APPLICANT: Hilburn, Erin  
 APPLICANT: Turner, C. Alexander Jr.  
 APPLICANT: Friedrich, Glenn  
 APPLICANT: Abuin, Alejandro  
 APPLICANT: Zambrowicz, Brian  
 APPLICANT: Sands, Arthur T.  
 APPLICANT: Walke, D. Wade  
 APPLICANT: Wilgowski, Nathaniel L.  
 APPLICANT: Kieke, James Alvin  
 APPLICANT: Potter, David George  
 TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
 TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
 FILE REFERENCE: LEX-0144-USA  
 CURRENT APPLICATION NUMBER: US/10/364,774  
 CURRENT FILING DATE: 2003-02-11  
 PRIOR APPLICATION NUMBER: US/09/795,926  
 PRIOR FILING DATE: 2001-02-28  
 PRIOR APPLICATION NUMBER: US 60/185,920  
 PRIOR FILING DATE: 2000-02-29  
 PRIOR APPLICATION NUMBER: US 60/186,553  
 PRIOR FILING DATE: 2000-03-02  
 PRIOR APPLICATION NUMBER: US 60/191,849

```

; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 27
; LENGTH: 269
; TYPE: PRT
; ORGANISM: homo sapiens
; US-10-364-774-27

Query Match 41.8%; Score 1370; DB 14; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e-124; Mismatches 0; Indels 0; Gaps 0;
Matches 256; Conservative 0; MisMatches 0; Length 269

Qy 98 MTDABEVDRYRENGFNIYVSDKLISLNRSLDPDHRHNCNSKRYLETLPNTSIIIPFHNGC 157
Db 1 MTDABEVDRYRENGFNIYVSDKLISLNRSLDPDHRHNCNSKRYLETLPNTSIIIPFHNGC 60
Qy 158 WSSLRILTIVSVLNSPPELVAEVILVDDSDREHKKPLDYLAFPSRLIRLTKREGI 217
Db 61 WSSLRILTIVSVLNSPPELVAEVILVDDSDREHKKPLDYLAFPSRLIRLTKREGI 120
Qy 218 IRTNLGASVATGDVITFLDSCHEANVNLPLDRIARNRKTVCPMDVTDIHDFFRE 277
Db 121 IRTNLGASVATGDVITFLDSCHEANVNLPLDRIARNRKTVCPMDVTDIHDFFRE 180
Qy 278 TQAGDAMRGAFDWEMYKRIPIPBLQKADSPDPRFSPYMAQSLIPAVDRKWFHLLGGDP 337
Db 181 TQAGDAMRGAFDWEMYKRIPIPBLQKADSPDPRFSPYMAQSLIPAVDRKWFHLLGGDP 240
Qy 338 GLRTWGGEQYELSPKV 353
Db 241 GLEIWGGEQYELSPKV 256

RESULT 15
US-09-795-926-35
Sequence 35, Application US/09795926
Patent No. US2002009486A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walko, D. Wade
; APPLICANT: Wilgowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kleke, James Alvin
; APPLICANT: Portee, David Alvin
; APPLICANT: George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
; CURRENT APPLICATION NUMBER: US/10/364,774
; CURRENT FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 35
; LENGTH: 321
; TYPE: PRT
; ORGANISM: homo sapiens
; US-10-364-774-35

Query Match 40.0%; Score 1311; DB 14; Length 321;
Best Local Similarity 99.2%; Pred. No. 1.2e-118; Mismatches 1; Indels 0; Gaps 0;
Matches 251; Conservative 1; MisMatches 1; Length 321

Qy 1 MRRKEKRLQIAVAVLVLNLVNGMVALYRERQDCTPGGSGAAVAPAGCOSHSHQK 60
Db 1 MRRKEKRLQIAVAVLVLNLVNGMVALYRERQDCTPGGSGAAVAPAGCOSHSHQK 60
Qy 61 KTFPLGDGOKLQKWHDKAIRDAQVNGEGRGPYMTDAEVYDQWRENGFNIYVSDK 120
Db 61 KTFPLGDGOKLQKWHDKAIRDAQVNGEGRGPYMTDAEVYDQWRENGFNIYVSDK 120
Qy 121 ISLRSLSLPDHRHNCNSKRYLETLPNTSIIIPFHNGSSLLRTVHSVLNSPPELVAI 180
Db 121 ISLRSLSLPDHRHNCNSKRYLETLPNTSIIIPFHNGSSLLRTVHSVLNSPPELVAI 180
Qy 181 VLVUDPSREHKKPLDYLAFPSRLIRLTKREGIIRTRMAGASVATGDVITFLDSC 240
Db 181 VLVUDPSREHKKPLDYLAFPSRLIRLTKREGIIRTRMAGASVATGDVITFLDSC 240
Qy 241 EANVNWLPPLGK 253
Db 241 EANVNWLPPLGK 253

RESULT 16
US-10-364-774-35
Sequence 35, Application US/10364774
; Publication No. US2003014497A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walko, D. Wade
; APPLICANT: Wilgowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kleke, James Alvin
; APPLICANT: Portee, David Alvin
; APPLICANT: George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
; CURRENT APPLICATION NUMBER: US/10/364,774
; CURRENT FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 35
; LENGTH: 321
; TYPE: PRT
; ORGANISM: homo sapiens
; US-10-364-774-35

Query Match 40.0%; Score 1311; DB 14; Length 321;
Best Local Similarity 99.2%; Pred. No. 1.2e-118; Mismatches 1; Indels 0; Gaps 0;
Matches 251; Conservative 1; MisMatches 1; Length 321

Qy 1 MRRKEKRLQIAVAVLVLNLVNGMVALYRERQDCTPGGSGAAVAPAGCOSHSHQK 60
Db 1 MRRKEKRLQIAVAVLVLNLVNGMVALYRERQDCTPGGSGAAVAPAGCOSHSHQK 60
Qy 61 KTFPLGDGOKLQKWHDKAIRDAQVNGEGRGPYMTDAEVYDQWRENGFNIYVSDK 120
Db 61 KTFPLGDGOKLQKWHDKAIRDAQVNGEGRGPYMTDAEVYDQWRENGFNIYVSDK 120
Qy 121 ISLRSLSLPDHRHNCNSKRYLETLPNTSIIIPFHNGSSLLRTVHSVLNSPPELVAI 180

```

RESULT 17  
US 09-925-301-1006  
Sequence 1006, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; CURRENT FILING DATE: 2001-08-10  
; PRIORITY APPLICATION NUMBER: PCT/US00/05882  
; PRIORITY FILING DATE: 2000-01-08  
; PRIORITY FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 29  
; SEQ ID NO: 1006  
; LENGTH: 561  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-925-301-1006

Query Match 34.3%; Score 1125; DB 9; Length 561;  
Best Local Similarity 44.0%; Pred. No. 3.9e-100; Indels 38; Gaps 16;  
Matches 232; Conservative 87; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGEGRPP--YPMTDAAERVQDQAYRENGFNYVSDKISLNRLSPDTRHNCNSKRYLETP 145  
Db 59 GPGGNGKPKPVVTPKEQKEMKMKFQINQFLNMASEMIALNRSPLDPLRLEGCKTKYQDNL 118

QY 146 NTSLIIPFNGEWSLLRTVHSVLNRSPELVAEVIVLUDFSREHUKPLEYM-ALFP 204  
Db 119 TTSVIVVFNEAWSTLRTVHSVINSPRMEEVILVLDASERDFKPLPESVKKV 178

QY 205 SVRITKKEGLTRMAGASVATGDTFLSHCEANVNLPLDPLRJARRKTCP 264  
Db 177 PVHVRMEORGSLGLRALKGAASVKGQQTFLDADCECTVGLWPLPLRARIKHDRTKTVCP 236

QY 265 MIDVHDPPRVEQAG-DAMRAFDWEMYKQIPIP--ELOKADSDPPEPSPWAGG 320  
Db 147 NTSLIIPFNGEWSLLRTVHSVLNRSPELVAEVIVLUDFSREHUKPLEYM-ALFP 204

QY 221 LFANDRKKWFLBLGGYDPPGLIEWGGEOYTSFKWMMGGERMEDTPCSRQGHIRKVKPVK 380  
Db 295 LFSIDRDPYQEGITVDAQMDWGENGLERFQIWCQGCTTLEVTCSSHVGHVFKATPPTF 354

QY 237 IIDVSDDPYB--MAGSDMTYGGFNWKLNFRMVPVOREMDRKGDKRMLPVRIPTMAGG 294  
Db 179 FVHVTRMEORGSLGLRALKGAQVITFLDAHCECTVGLWPLDPLRJARRKTCP 238

QY 265 MIDVHDPPRVEQAG-DAMRAFDWEMYKQIPIP--ELOKADSDPPEPSPWAGG 320  
Db 239 IDIVSDDPYB--MAGSDMTYGGFNWKLNFRMVPVOREMDRKGDKRMLPVRIPTMAGG 295

QY 321 LPAVDRKWWFLBLGGYDPPGLIEWGGEOYTSFKWMMGGERMEDTPCSRQGHIRKVKPVK 380  
Db 297 LFSIDRDPYQEGITVDAQMDWGENGLERFQIWCQGCTTLEVTCSSHVGHVFKATPPTF 354

QY 381 PAGVS--LARNLKVRAEVWDEAYTIVQRPYRHLASCTDVAVKKLSSLNQSKMF 438  
Db 355 PGGTQQININKRNLRAEVWDEFKFFPPIISPQVTKVYDGSRLGRLKQCRPFWSY 414

QY 439 MTKLAWD--LPKEYPPVPEPPAAMWGEINVGTGICADDTHGALGSPPLLEGCVRGGEAA 496  
Db 321 LPAVDRKWWFLBLGGYDPPGLIEWGGEOYTSFKWMMGGERMEDTPCSRQGHIRKVKPVK 380  
Db 381 PAGVS--LARNLKVRAEVWDEAYTIVQRPYRHLASCTDVAVKKLSSLNQSKMF 438  
QY 357 PGGTQOININKRNLRAEVWDEFKFFPPIISPQVTKVYDGSRLGRLKQCRPFWSY 416  
QY 439 MTKLAWD--LPKEYPPVPEPPAAMWGEINVGTGICADDTHGALGSPPLLEGCVRGGEAA 496  
Db 417 LENTYPSQIPRY-----SSLGEBIRNVTENOCLDRNMARKENKEVQGJFNC-HGMG--- 465

QY 497 WNNQVFTTWRERDIPGDQHPTKKPQFDASHTSPVLTBDSKMSKGQWLYKRDK-TL 555  
Db 466 --GNQVSYTANKEIRTDD-----LCLDVSKLNGPVIMKCHLKGQWLYDVKLT 517

QY 556 YHPVSGSCMD-CSESQHRI-FMNTCNPSSSTQOQWLFHENTMSTVKEP 600  
Db 518 QHVNNSNQCLDKATBEDSQQPSIRDCS-GSRSQQWLLR--NVTLPBIF 559

RESULT 18  
US 10-001-851-24  
Sequence 24, Application US/10001851  
; Publication No. US20020115628A1  
; GENERAL INFORMATION:  
; APPLICANT: MEYERS, Rachel A.  
; TITLE OF INVENTION: 47169 and 33935, NO. US20020115628A1el Human Glycosyl Transferase  
; FILE REFERENCE: 1047-5601  
; CURRENT APPLICATION NUMBER: US/10/001,851  
; CURRENT FILING DATE: 2001-11-20  
; PRIORITY APPLICATION NUMBER: US 60/249,939  
; PRIORITY FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 29  
; SEQ ID NO: 24  
; LENGTH: 559  
; TYPE: PRT  
; ORGANISM: Sub sp.  
; US-10-001-851-24

Query Match 34.1%; Score 1117; DB 13; Length 559;  
Best Local Similarity 43.6%; Pred. No. 2.4e-99; Indels 38; Gaps 16;  
Matches 230; Conservative 90; Mismatches 169; Indels 38; Gaps 16;

QY 57 GPGGNGKPKPVVTPKEQKEMKMKFQINQFLNMASEMIALNRSPLDPLRLEGCKTKYQDNL 116  
Db 146 NTSLIIPFNGEWSLLRTVHSVLNRSPELVAEVIVLUDFSREHUKPLEYM-ALFP 204  
QY 177 PVHVRMEORGSLGLRALKGAASVKGQQTFLDADCECTVGLWPLPLRARIKHDRTKTVCP 236  
Db 205 SVRITKKEGLTRMAGASVATGDTFLSHCEANVNLPLDPLRJARRKTCP 264  
QY 265 MIDVHDPPRVEQAG-DAMRAFDWEMYKQIPIP--ELOKADSDPPEPSPWAGG 320  
Db 221 LPAVDRKWWFLBLGGYDPPGLIEWGGEOYTSFKWMMGGERMEDTPCSRQGHIRKVKPVK 380  
QY 281 PAGVS--LARNLKVRAEVWDEAYTIVQRPYRHLASCTDVAVKKLSSLNQSKMF 438  
Db 355 PGGTQQININKRNLRAEVWDEFKFFPPIISPQVTKVYDGSRLGRLKQCRPFWSY 414  
QY 439 MTKLAWD--LPKEYPPVPEPPAAMWGEINVGTGICADDTHGALGSPPLLEGCVRGGEAA 496  
Db 415 LENTYPSQIPRY-----SSLGEBIRNVTENOCLDRNMARKENKEVQGJFNC-HGMG--- 463  
QY 497 WNNQVFTTWRERDIPGDQHPTKKPQFDASHTSPVLTBDSKMSKGQWLYKRDK-TL 555  
Db 464 --GNQVSYTANKEIRTDD-----LCLDVSKLNGPVIMKCHLKGQWLYDVKLT 515  
QY 556 YHPVSGSCMD-CSESQHRI-FMNTCNPSSSTQOQWLFHENTMSTVKEP 600  
Db 516 QHVNNSNQCLDKATBEDSQQPSIRDCS-GSRSQQWLLR--NVTLPBIF 559

RESULT 19  
US 10-001-851-21  
Sequence 21, Application US/10001851  
; Publication No. US20020115628A1  
; GENERAL INFORMATION:  
; APPLICANT: MEYERS, Rachel A.  
; TITLE OF INVENTION: 47169 and 33935, NO. US20020115628A1el Human Glycosyl Transferase

; TITLE OF INVENTION: Uses Thereof  
 FILE REFERENCE: 10147-56UL  
 CURRENT APPLICATION NUMBER: US/10/001,851  
 CURRENT FILING DATE: 2001-11-20  
 PRIORITY APPLICATION NUMBER: US 60/249,939  
 PRIORITY FILING DATE: 2000-11-20  
 NUMBER OF SEQ ID NOS: 29  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO: 21  
 LENGTH: 559  
 TYPE: PRT  
 ORGANISM: Rattus sp.  
 ; OTHER INFORMATION: Polypeptide GalNAc transferase  
 US-10-001-851-21  
 Query Match 34.0%; Score 1116; DB 13; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 2.9e-99;  
 Matches 230; Conservative 88; Mismatches 171; Indels 38; Gaps 16;  
 Qy 88 GNGEQQRP--YMTDAERVDQAYRENGFNIYVSDKISLNRSIDPDIRHPCNSKRYLTLP 145  
 Db 57 GPGEMSKPWPVPKEDQEKMKPQKINQFNLMASEMIAFNRSIDPDIRLEGCKTVYVYDLP 116  
 Qy 146 NSIIITPPHNGWSSLRTVSVLNSRPELVAETVLTVDSDRENLKKPFDYMLALFP 204  
 Db 117 TTSVWVVFHNKEAWSTLRTVSVLNSRPELVAETVLTVDSDRENLKKPFDYMLALFP 204  
 Qy 265 MIDVHDDFRYETQAG-DAMRGAFTWEMYKRIPIP--ELQKADPSDFPESPMAGG 320  
 Db 237 IIDVISDDTFEY--MAGSDMTYGGFNFNKLWPRWVPPQREMDRRKGDRTLVPTPTMAGG 294  
 Qy 321 LFAVDKRWFWLGGYGPGLETWGGEYEISPKWMMGGRMEDIPCSRQGHYTKVYKV 380  
 Db 295 LFSIDRDYFQETGTYAGMDTWGGENLNEISFPIWIQCGTETLVTCTCERGHYFRTKATPYF 354  
 Qy 381 PAGVS--LARNUKRVAEVWMDYEAYIYORRPEYRHLASAGDVAVQKKLRSLSNCKSFKWF 438  
 Db 355 PGGTGGQIINKNRRRLAEMWMDPEFKNFFYIISPGVTVKDYGDSRGLRHKLOCKPSWY 414  
 Qy 439 MTKIAND--LPKPYPPVEPPAAWGEIRNVGJGLCADTKHGAIGSPURLEGCVRGGEAA 496  
 Db 415 LENIYDPSQIPHY-----FSLGEIRNVETNQCDNMARKENERKVGFNC-HGMG-- 463  
 Qy 497 WNMNOVFTTWRDTRPGDPOHTKKCFDAISHTSPVTLCHSMKGQNLWQKRYKOK-TL 555  
 Db 464 --GNQDPSYTAKEIRDD-----LCLDVSKLNGPVTMVKCQHKGQNLWQYDPVQLTL 515  
 Qy 556 YHPVSSCMD-CSESDRHRI-FMNTCNPSSLTQMLFEHTNSTVLEKP 600  
 Db 516 OHVNSNQCLDKRATERBDSQVSPSRDCT-GSRSSQWLLR--NVTLPETP 559  
 RESULT 20  
 US-10-205-219-36  
 ; Sequence 36, Application US/10205219  
 ; Publication No. US20030138803A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Warner-Lambert Company  
 ; APPLICANT: Lee, Kevin  
 ; APPLICANT: Dixon, Alastair  
 ; APPLICANT: Brookbank, Robert  
 ; APPLICANT: Pincock, Robert  
 ; APPLICANT: Warner-Lambert Company  
 ; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain  
 ; FILE REFERENCE: WIL-A-018200  
 ; CURRENT APPLICATION NUMBER: US/10/205,219  
 ; CURRENT FILING DATE: 2002-07-24  
 ; PRIORITY APPLICATION NUMBER: GB 0118354.0  
 ; PRIORITY FILING DATE: 2001-07-27  
 ; NUMBER OF SEQ ID NOS: 197  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO: 76  
 LENGTH: 559  
 TYPE: PRT  
 ORGANISM: Rattus norvegicus  
 ; OTHER INFORMATION: Polypeptide GalNAc transferase  
 US-10-205-219-36  
 Query Match 34.0%; Score 1116; DB 14; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 2.9e-99;  
 Matches 230; Conservative 88; Mismatches 171; Indels 38; Gaps 16;  
 Qy 88 GNGEQQRP--YMTDAERVDQAYRENGFNIYVSDKISLNRSIDPDIRHPCNSKRYLTLP 145  
 Db 57 GPGEMSKPWPVPKEDQEKMKPQKINQFNLMASEMIAFNRSIDPDIRLEGCKTVYVYDLP 116  
 Qy 146 NSIIITPPHNGWSSLRTVSVLNSRPELVAETVLTVDSDRENLKKPFDYMLALFP 204  
 Db 117 TTSVWVVFHNKEAWSTLRTVSVLNSRPELVAETVLTVDSDRENLKKPFDYMLALFP 204  
 Qy 265 MIDVHDDFRYETQAG-DAMRGAFTWEMYKRIPIP--ELQKADPSDFPESPMAGG 320  
 Db 237 IIDVISDDTFEY--MAGSDMTYGGFNFNKLWPRWVPPQREMDRRKGDRTLVPTPTMAGG 294  
 Qy 321 LFAVDKRWFWLGGYGPGLETWGGEYEISPKWMMGGRMEDIPCSRQGHYTKVYKV 380  
 Db 295 LFSIDRDYFQETGTYAGMDTWGGENLNEISFPIWIQCGTETLVTCTCERGHYFRTKATPYF 354  
 Qy 381 PAGVS--LARNUKRVAEVWMDYEAYIYORRPEYRHLASAGDVAVQKKLRSLSNCKSFKWF 438  
 Db 355 PGGTGGQIINKNRRRLAEMWMDPEFKNFFYIISPGVTVKDYGDSRGLRHKLOCKPSWY 414  
 Qy 439 MTKIAND--LPKPYPPVEPPAAWGEIRNVGJGLCADTKHGAIGSPURLEGCVRGGEAA 496  
 Db 415 LENIYDPSQIPHY-----FSLGEIRNVETNQCDNMARKENERKVGFNC-HGMG-- 463  
 Qy 497 WNMNOVFTTWRDTRPGDPOHTKKCFDAISHTSPVTLCHSMKGQNLWQKRYKOK-TL 555  
 Db 464 --GNQDPSYTAKEIRDD-----LCLDVSKLNGPVTMVKCQHKGQNLWQYDPVQLTL 515  
 Qy 556 YHPVSSCMD-CSESDRHRI-FMNTCNPSSLTQMLFEHTNSTVLEKP 600  
 Db 516 OHVNSNQCLDKRATERBDSQVSPSRDCT-GSRSSQWLLR--NVTLPETP 559  
 RESULT 21  
 US-10-205-219-76  
 ; Sequence 76, Application US/10205219  
 ; Publication No. US20030138803A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Warner-Lambert Company  
 ; APPLICANT: Lee, Kevin  
 ; APPLICANT: Dixon, Alastair  
 ; APPLICANT: Brookbank, Robert  
 ; APPLICANT: Pincock, Robert  
 ; APPLICANT: Warner-Lambert Company  
 ; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain  
 ; FILE REFERENCE: WIL-A-018200  
 ; CURRENT APPLICATION NUMBER: US/10/205,219  
 ; CURRENT FILING DATE: 2002-07-24  
 ; PRIORITY APPLICATION NUMBER: GB 0118354.0  
 ; PRIORITY FILING DATE: 2001-07-27  
 ; NUMBER OF SEQ ID NOS: 197  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO: 76  
 LENGTH: 559  
 TYPE: PRT  
 ORGANISM: Rattus norvegicus  
 ; OTHER INFORMATION: Polypeptide GalNAc transferase

US-10-205-219-76

Query Match 34.0%; Score 1116; DB 14; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 2.9e-99;  
 Matches 230; Conservative 88; Mismatches 171; Indels 38; Gaps 16;

Qy 88 GNGFOGRR--YPMTDAAEVDQAYRENGENYVSDKISLRSRSLPDRHNCNSKYLETP 145  
 Db 57 GPGEMGKPVVTPKEDOKEKMKEMKINQFLNMASEMIALNNSLPDRLEGCKTKYDPDLP 115

Qy 146 NTSTIIPNEGNSLLRTVHSYVLSRSPELVAEVLVDDFSRDLRHKPLEDM-ALFP 204  
 Db 117 TTSVIVVFNARNEANSTLRTVHSVNRSPRHMELBLVLDASERDFLRKPLESYVKKV 176

Qy 205 SVRLRTKREGLRTMIGASVATGVDITFLDSHCEANVNPPLDRIARRKTVCP 264  
 Db 177 PURVTRMSEGRSLRARLKGAAVSRGOTITFLDAHCETAGMLBLPLRARIKHDRTVCP 236

Qy 265 MIDVHDHPRYETQAG-DAMRGAFDWMYKRPPIP--ELQADSDPFPSPVMGG 320  
 Db 237 IDVSDRFBY-MAGSDMTYGGFNMUNFRWYVPUQREMDRKGDTLPVPTMAGG 294

Qy 295 LFSIDRDFQBDIGTYDAMIDWGGENLTSFRIWCGGTLETIVTCSHVGHVERKATPTF 354

Qy 321 LFAVDRKWWLGEQDPLBLWGGQYELSFKVMGGMEDIFCSRVGHIYRKVPIV 380

Qy 381 PAGVS-LARNLKVRAEVMDDEVAYIQRPRPYRHLASAGDVAVKURSSLNCKSKFV 438

Db 355 PGSGQOINNNRRLAEVWNDERKNNFFYIISPGVTKYDQGKPSWY 414

Qy 397 WNNQMVFTFWREDTRPGDPQHTKCFDAISHTSPVTLYDCHSKNGNDLWKRKD-TL 555

Db 439 MTKIAWD--LPKFVPPVSPPAAMGEIRNVTGICADTTKHALGSPPLBLGCYRGGRAA 496

Qy 464 --GNQVFSYANKERITD----LCDVSKLNGPVMLKCHHLKGQNLWEIDPVKLT 515

Db 415 LENYVPSQIPIV-----FSLGEIRVNETQOCLDNMARKENEKJIFNC-HGMG-- 463

Qy 497 WNNQMVFTFWREDTRPGDPQHTKCFDAISHTSPVTLYDCHSKNGNDLWKRKD-TL 555

Db 464 --GNQVFSYANKERITD----LCDVSKLNGPVMLKCHHLKGQNLWEIDPVKLT 515

Qy 556 YHPVSGSCMD-CSSDHRI-FMNTCNPSSLTQWIFEHINSTVLEKP 600

Db 556 YHPVSGSCMD-CSSDHRI-FMNTCNPSSLQWIFEHINSTVLEKP 600

Qy 516 QHVNNSNQCLDKATEEDSQVPSIRD-GSRQQMLR--NVTLPBIF 559

Db 516 QHVNNSNQCLDKATEEDSQVPSIRD-GSRQQMLR--NVTLPBIF 559

RESULT 22

US-10-001-851-22

; Sequence 22, Application US/10001851

; Publication No. US20020115628A1

; GENERAL INFORMATION:

; APPLICANT: MEYERS, Rachel A.

; TITLE OF INVENTION: 47169 and 33935, NO. US20020115628A1el Human Glycosyl Transferase

; FILE REFERENCE: 10147-5601

; CURRENT APPLICATION NUMBER: US/10/001,851

; CURRENT FILING DATE: 2001-11-20

; PRIOR APPLICATION NUMBER: US 60/249,939

; PRIOR FILING DATE: 2000-11-20

; NUMBER OF SEQ ID NOS: 29

; NUMBER OF SEQ ID NOS: 29

; APPLICANT: MEYERS, Rachel A.

; TITLE OF INVENTION: 47169 and 33935, NO. US20020115628A1el Human Glycosyl Transferase

; FILE REFERENCE: 10147-5601

; CURRENT APPLICATION NUMBER: US/10/001,851

; CURRENT FILING DATE: 2001-11-20

; PRIOR APPLICATION NUMBER: US 60/249,939

; PRIOR FILING DATE: 2000-11-20

; NUMBER OF SEQ ID NOS: 29

; NUMBER OF SEQ ID NOS: 29

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO: 23

; LENGTH: 559

; TYPE: PRT

; ORGANISM: BOB sp.

; US-10-001-851-23

Query Match 34.0%; Score 1115; DB 13; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 3.7e-99;  
 Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

Qy 88 GNGFOGRR--YPMTDAAEVDQAYRENGENYVSDKISLRSRSLPDRHNCNSKYLETP 145  
 Db 57 GPGEMGKPVVTPKEDOKEKMKEMKINQFLNMASEMIALNNSLPDRLEGCKTKYDPDLP 116

Qy 146 NTSTIIPNEGNSLLRTVHSYVLSRSPELVAEVLVDDFSRDLRHKPLEDM-ALFP 204  
 Db 117 TTSVIVVFNARNEANSTLRTVHSVNRSPRHMELBLVLDASERDFLRKPLESYVKKV 176

Qy 205 SVRLRTKREGLRTMIGASVATGVDITFLDSHCEANVNPPLDRIARRKTVCP 264  
 Db 177 PVHVRMEORSGLRLARLKGAAVSKGQVTFLDACETCIVGWLPLRARIKHDRTVCP 236

Qy 265 MIDVHDHPRYETQAG-DAMRGAFDWMYKRPPIP--ELQADSDPFPSPVMGG 320

RESULT 23

US-10-001-851-23

; Sequence 23, Application US/10001851

; Publication No. US20020115628A1

; GENERAL INFORMATION:

; APPLICANT: MEYERS, Rachel A.

; TITLE OF INVENTION: 47169 and 33935, NO. US20020115628A1el Human Glycosyl Transferase

; FILE REFERENCE: 10147-5601

; CURRENT APPLICATION NUMBER: US/10/001,851

; CURRENT FILING DATE: 2001-11-20

; PRIOR APPLICATION NUMBER: US 60/249,939

; PRIOR FILING DATE: 2000-11-20

; NUMBER OF SEQ ID NOS: 29

; NUMBER OF SEQ ID NOS: 29

; APPLICANT: MEYERS, Rachel A.

; TITLE OF INVENTION: 47169 and 33935, NO. US20020115628A1el Human Glycosyl Transferase

; FILE REFERENCE: 10147-5601

; CURRENT APPLICATION NUMBER: US/10/001,851

; CURRENT FILING DATE: 2001-11-20

; PRIOR APPLICATION NUMBER: US 60/249,939

; PRIOR FILING DATE: 2000-11-20

; NUMBER OF SEQ ID NOS: 29

; NUMBER OF SEQ ID NOS: 29

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO: 23

; LENGTH: 559

; TYPE: PRT

; ORGANISM: BOB sp.

; US-10-001-851-23

Query Match 34.0%; Score 1115; DB 13; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 3.7e-99;  
 Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

Qy 88 GNGFOGRR--YPMTDAAEVDQAYRENGENYVSDKISLRSRSLPDRHNCNSKYLETP 145  
 Db 57 GPGEMGKPVVTPKEDOKEKMKEMKINQFLNMASEMIALNNSLPDRLEGCKTKYDPDLP 116

Qy 146 NTSTIIPNEGNSLLRTVHSYVLSRSPELVAEVLVDDFSRDLRHKPLEDM-ALFP 204  
 Db 117 TTSVIVVFNARNEANSTLRTVHSVNRSPRHMELBLVLDASERDFLRKPLESYVKKV 176

Qy 205 SVRLRTKREGLRTMIGASVATGVDITFLDSHCEANVNPPLDRIARRKTVCP 264  
 Db 177 PVHVRMEORSGLRLARLKGAAVSKGQVTFLDACETCIVGWLPLRARIKHDRTVCP 236

Qy 265 MIDVHDHPRYETQAG-DAMRGAFDWMYKRPPIP--ELQADSDPFPSPVMGG 320

Db 237 IIVDVISDDTFEY--MAGSDMTYGGFNWKLNFRWYFVPOREMDRKGDRTLPVRTPTMAGG 294  
 Qy 321 LFAUDRKWFWELGYDPGLIWMGQEVTSRKYWMCGGRMEDIPCSRUYHVKVVKV 380  
 Db 295 LFSIDRDYFQEIGTYDAGMDIWGSGENLBSRPTWQCGSTLEITVCSHGVFRKATPYF 354  
 Qy 381 PAGVS--LARNLKRVAEVMDAYEAYIYKORRPEVRHLSAGDVAQKURSSLNCKSKWV 438  
 Db 355 PGGQQLINKNKRNLAEVMDEFFQNFVYIISPGVTKVQDGFISRSRGLRKLQCRPSWY 414  
 Qy 439 MTKIAWD-LPKFVPPVPEPPAAMGBERINGTGLCADTRKGAGSGSPRLLEGCVTRGRGAA 436  
 Db 415 LENIYPSDQIPRHY-----FSLGBIRNVTENOCLDNMARKENEBKVGFNC-HGMC--- 463  
 Qy 497 WNNMQVFTTWRERTRPGPQRTKFCFAISHTSPVTLYDCHSMKGNLWVKYRDK-TL 555  
 Db 464 --GNQVFSTANKERBIRTD-----LCIDVSKUNGPMVILKCHLKGNLWEDPVKJTL 515  
 Qy 556 YHPVSGSCMD-CBSDHRI-PMNTCNPSLTLQWLFETNSTLEKF 600  
 Db 516 OHVNSNQCLDKATBEDSQVPSIRDNC-GSRSQWLLR--NVTUPEIF 559

RESULT 24  
 US-10-001-851-20  
 ; Sequence 20, Application US/10001851  
 ; Publication No. US20020115628A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MEYERS, Rachel A.  
 ; APPLICANT: WILLIAMSON, Mark  
 ; TITLE OF INVENTION: Uses Thereof  
 ; TITLE OF INVENTION: USES THEREOF  
 ; FILE REFERENCE: 10147-5601  
 ; CURRENT APPLICATION NUMBER: US/10/001,851  
 ; CURRENT FILING DATE: 2001-11-20  
 ; PRIOR APPLICATION NUMBER: US 60/249,939  
 ; PRIOR FILING DATE: 2000-11-20  
 ; NUMBER OF SEQ ID NOS: 29  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 20  
 ; LENGTH: 559  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-001-851-20

Query Match 34.0%; Score 1113; DB 13; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 5.8e-99; Mismatches 16;  
 Matches 230; Conservative 87; Mismatches 172; Indels 38; Gaps 16;

Qy 88 GNGBQGRP--YPMDAERYDQAYRENGENIYVSKISLNRSPDPRHENCNKRYLETP 145  
 Db 57 GPGEWKGPKVVKIPKEDQEKMKMENQNLMASENMLRSPDVRLEGCKTKVDPNLP 116

Qy 145 NTSLIIPFNEGMSLRTVHSVLRSPRELVARIVDPSREHJKPLEYM-ALFP 204  
 Db 117 TSVVIVFVNEANSTLRTVHSVLRSPRELVARIVDPSREHJKPLEYM-AFP 176

Qy 205 SVRLRTKREGLRTTRMGASVATGDTVFLDSCHEANVNPPLIARARRKTVCP 264  
 Db 177 PVPVIRMRSGLRARLGAVKQVITFLDCECTVWMLPPLARIKHRRTVCP 236

Qy 265 MIDVIDHDFRFEYTOAG-DAMGRADFDMEMYKRIPIP--ELOKADPSDPFSPWMAGG 320  
 Db 237 IDVISDDTFEY--MAGSDMTYGGFNWKLNFRWYFVPOREMDRKGDRTLPVRTPTMAGG 294

Qy 321 LFAUDRKWFWELGYDPGLIWMGQEVTSRKYWMCGGRMEDIPCSRUYHVKVVKV 380  
 Db 294 LFSIDRDYFQEIGTYDAGMDIWGSGENLBSRPTWQCGSTLEITVCSHGVFRKATPYF 353

Qy 381 PAGVS--LARNLKRVAEVMDAYEAYIYKORRPEVRHLSAGDVAQKURSSLNCKSKWV 438  
 Db 354 PGGQQLINKNKRNLAEVMDEFFQNFVYIISPGVTKVQDGFISRSRGLRKLQCRPSWY 413  
 Qy 439 MTKIAWD-LPKFVPPVPEPPAAMGBERINGTGLCADTRKGAGSGSPRLLEGCVTRGRGAA 436  
 Db 414 LENI-----YPSDQIPRHYVSLGEIRNVTENOCLDNMARKENEBKVGFNC-HGMC--- 462

RESULT 25  
 US-10-292-856-62  
 ; Sequence 62, Application US/10292896  
 ; Publication No. US2003016850A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HASSAN, Helle  
 ; APPLICANT: REIS, Celso A.  
 ; APPLICANT: BENNETT, Eric P.  
 ; APPLICANT: CLAUSEN, Henrik  
 ; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GAMES  
 ; TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS  
 ; FILE REFERENCE: 4305/1H154-US33  
 ; CURRENT APPLICATION NUMBER: US/10/292,896  
 ; CURRENT FILING DATE: 2003-03-17  
 ; PRIOR APPLICATION NUMBER: US 60/425,204  
 ; PRIOR FILING DATE: 2002-11-09  
 ; PRIOR APPLICATION NUMBER: PCT/DK01/00328  
 ; PRIOR FILING DATE: 2001-05-10  
 ; PRIOR APPLICATION NUMBER: US 60/203,331  
 ; PRIOR FILING DATE: 2000-05-11  
 ; NUMBER OF SEQ ID NOS: 130  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 62  
 ; LENGTH: 556  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-292-856-62

Query Match 33.7%; Score 1106; DB 14; Length 556;  
 Best Local Similarity 43.8%; Pred. No. 2.8e-98; Mismatches 165; Indels 36; Gaps 15;  
 Matches 227; Conservative 90; Mismatches 165; Indels 36; Gaps 15;

Qy 88 GNGBQGRP--YPMDAERYDQAYRENGENIYVSKISLNRSPDPRHENCNKRYLETP 145  
 Db 56 GPGEWKGPKVVKIPKEDQEKMKMENQNLMASENMLRSPDVRLEGCKTKVDPNLP 115

Qy 146 NTSLIIPFNEGMSLRTVHSVLRSPRELVARIVDPSREHJKPLEYM-AFP 204  
 Db 116 NTSLIIPFNEGMSLRTVHSVLRSPRELVARIVDPSREHJKPLEYM-AFP 176

Qy 205 SVRLRTKREGLRTTRMGASVATGDTVFLDSCHEANVNPPLIARARRKTVCP 264  
 Db 176 PVPVIRMRSGLRARLGAVKQVITFLDCECTVWMLPPLARIKHRRTVCP 235

Qy 265 MIDVIDHDFRFEYTOAG-DAMGRADFDMEMYKRIPIP--ELOKADPSDPFSPWMAGG 320  
 Db 236 IDVISDDTFEY--MAGSDMTYGGFNWKLNFRWYFVPOREMDRKGDRTLPVRTPTMAGG 293

Qy 321 LFAUDRKWFWELGYDPGLIWMGQEVTSRKYWMCGGRMEDIPCSRUYHVKVVKV 380  
 Db 294 LFSIDRDYFQEIGTYDAGMDIWGSGENLBSRPTWQCGSTLEITVCSHGVFRKATPYF 353

Qy 381 PAGVS--LARNLKRVAEVMDAYEAYIYKORRPEVRHLSAGDVAQKURSSLNCKSKWV 438  
 Db 354 PGGQQLINKNKRNLAEVMDEFFQNFVYIISPGVTKVQDGFISRSRGLRKLQCRPSWY 413  
 Qy 439 MTKIAWD-LPKFVPPVPEPPAAMGBERINGTGLCADTRKGAGSGSPRLLEGCVTRGRGAA 436  
 Db 414 LENI-----YPSDQIPRHYVSLGEIRNVTENOCLDNMARKENEBKVGFNC-HGMC--- 462

QY 497 WNNMQVFTTWREDIRPGDPOQHTKRCFDAISHTSPVTLYDCHSNKGNOIWKYRKDK-TL 555  
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
Db 463 -GNQVFSYRDAKEIRDD----LCLDVSRINGPVIMKCHMRGNOLWEYDAERLT 514  
QY 556 YHPVSGSCMD-CSESDHRI-FMNTCNPSSLTQWIFEH 591  
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
Db 515 RHVNSNQCLDEPSEEDKMQVTPMQDCS-GSRSQWMLRN 551

Search completed: November 22, 2004, 13:42:20  
Job time : 58 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using SW model

Run on: November 22, 2004, 13:38:52 ; Search time: 21 seconds  
(without alignments)  
2762.796 Million cell updates/sec

## Title: US-10-001-851-2

## Perfect score: 3278

## Sequence: 1 MRRKEKRLIQAVALVLAALV.....TQQLFERTNSTVLEKFNEN 603

## Scoring table: BLOSUM62

Gapop 10.0 , Gapext: 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : PIR\_79;\*

1: pir1;\*  
2: pir2;\*  
3: pir3;\*  
4: pir4;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	1392.5	42.5	684	2	T26930		hypothetical protein
2	1392	42.5	276	2	T12525		hypothetical protein
3	1125	34.3	559	2	JC4223		polypeptide N-acetyl
4	1113	34.0	559	2	AA5987		polypeptide N-acetyl
5	1113	34.0	601	2	T42521		polypeptide N-acetyl
6	1083	33.0	624	2	T42247		polypeptide N-acetyl
7	1079	32.9	626	2	T42246		polypeptide N-acetyl
8	1075.5	32.9	623	2	T42245		probable polypeptide
9	988.5	30.2	563	2	A88315		probable polypeptide N-acetyl
10	988.5	30.2	612	2	T42243		probable polypeptide N-acetyl
11	968	29.5	617	2	T42249		probable polypeptide N-acetyl
12	949	29.0	571	2	T37405		probable polypeptide N-acetyl
13	948.5	28.9	618	2	T42248		probable polypeptide N-acetyl
14	938	28.6	562	2	T42250		probable polypeptide N-acetyl
15	934	28.5	633	2	JC5247		probable polypeptide N-acetyl
16	930	27.5	579	2	T31549		probable polypeptide N-acetyl
17	883	26.9	589	2	T42244		probable polypeptide N-acetyl
18	801	24.4	605	2	T27397		hypothetical protein
19	632	19.3	421	2	T42252		hypothetical protein
20	147	4.5	1044	2	H97186		hypothetical protein
21	139	4.2	306	2	D89531		hypothetical protein
22	138	4.2	308	2	H89506		hypothetical protein
23	136	4.1	328	2	B84263		buccinoglycan biosynthesis
24	135.5	4.1	312	2	S74659		hypothetical protein
25	135	4.1	318	2	D87506		glycosyl transferase
26	130	4.0	322	2	AC2023		hypothetical protein
27	130	4.0	787	2	H98163		hypothetical protein
28	128	3.9	470	2	C7641		hypothetical protein
29	126	3.8	313	2	AI2404		hypothetical protein
30							dolichol-phosphate
31							glucuronosyl transferase
32							hypothetical protein
33							UDP-hexose transferase
34							UDP-hexose transferase
35							probable glycosyl transferase
36							glycosyl transferase
37							conserved hypothetical protein
38							glycosyl transferase
39							probable glucosidase
40							glycosyl transferase
41							glycosyl transferase
42							hypothetical protein
43							glycosyl transferase
44							hypothetical protein
45							endo-1,4-beta-D-xylose isomerase
46							hypothetical protein
47							glycosyl transferase
48							dolichyl-phosphate
49							glycosyl transferase
50							hypothetical protein
51							xylose isomerase
52							microtubule-associated protein
53							hypothetical protein
54							probable glycosyl transferase
55							probable dolichol-phosphate
56							hypothetical protein
57							glycosyl transferase
58							dolichol-P-glucosidase
59							aspartate-tRNA ligase
60							phosphoribosylformylglutamate synthase
61							dolichol-phosphate
62							probable glycosyl transferase
63							glycosyl transferase
64							exoco protein - Rho
65							cellulose synthase
66							galactosidase I precursor
67							galactosidase I precursor
68							hypothetical protein
69							glycosyl transferase
70							hypothetical protein
71							galactosidase I precursor
72							galactosidase I precursor
73							hypothetical protein
74							hypothetical protein
75							glycosyl transferase
76							hypothetical protein
77							nitrogenase molybdate protein
78							nitrogenase molybdate protein
79							nitrogenase molybdate protein
80							nitrogenase molybdate protein
81							nitrogenase molybdate protein
82							nitrogenase molybdate protein
83							nitrogenase molybdate protein
84							nitrogenase molybdate protein
85							nitrogenase molybdate protein
86							nitrogenase molybdate protein
87							nitrogenase molybdate protein
88							nitrogenase molybdate protein
89							nitrogenase molybdate protein
90							nitrogenase molybdate protein
91							nitrogenase molybdate protein
92							nitrogenase molybdate protein
93							nitrogenase molybdate protein
94							nitrogenase molybdate protein
95							nitrogenase molybdate protein
96							nitrogenase molybdate protein
97							nitrogenase molybdate protein
98							nitrogenase molybdate protein
99							nitrogenase molybdate protein
100							nitrogenase molybdate protein
101							nitrogenase molybdate protein
102							nitrogenase molybdate protein
103							nitrogenase molybdate protein
104							nitrogenase molybdate protein
105							nitrogenase molybdate protein
106							nitrogenase molybdate protein
107							nitrogenase molybdate protein
108							nitrogenase molybdate protein
109							nitrogenase molybdate protein
110							nitrogenase molybdate protein
111							nitrogenase molybdate protein
112							nitrogenase molybdate protein
113							nitrogenase molybdate protein
114							nitrogenase molybdate protein
115							nitrogenase molybdate protein
116							nitrogenase molybdate protein
117							nitrogenase molybdate protein
118							nitrogenase molybdate protein
119							nitrogenase molybdate protein
120							nitrogenase molybdate protein
121							nitrogenase molybdate protein
122							nitrogenase molybdate protein
123							nitrogenase molybdate protein
124							nitrogenase molybdate protein
125							nitrogenase molybdate protein
126							nitrogenase molybdate protein
127							nitrogenase molybdate protein
128							nitrogenase molybdate protein
129							nitrogenase molybdate protein
130							nitrogenase molybdate protein
131							nitrogenase molybdate protein
132							nitrogenase molybdate protein
133							nitrogenase molybdate protein
134							nitrogenase molybdate protein
135							nitrogenase molybdate protein
136							nitrogenase molybdate protein
137							nitrogenase molybdate protein
138							nitrogenase molybdate protein
139							nitrogenase molybdate protein
140							nitrogenase molybdate protein
141							nitrogenase molybdate protein
142							nitrogenase molybdate protein
143							nitrogenase molybdate protein
144							nitrogenase molybdate protein
145							nitrogenase molybdate protein
146							nitrogenase molybdate protein
147							nitrogenase molybdate protein
148							nitrogenase molybdate protein
149							nitrogenase molybdate protein
150							nitrogenase molybdate protein
151							nitrogenase molybdate protein
152							nitrogenase molybdate protein
153							nitrogenase molybdate protein
154							nitrogenase molybdate protein
155							nitrogenase molybdate protein
156							nitrogenase molybdate protein
157							nitrogenase molybdate protein
158							nitrogenase molybdate protein
159							nitrogenase molybdate protein
160							nitrogenase molybdate protein
161							nitrogenase molybdate protein
162							nitrogenase molybdate protein
163							nitrogenase molybdate protein
164							nitrogenase molybdate protein
165							nitrogenase molybdate protein
166							nitrogenase molybdate protein
167							nitrogenase molybdate protein
168							nitrogenase molybdate protein
169							nitrogenase molybdate protein
170							nitrogenase molybdate protein
171							nitrogenase molybdate protein
172							nitrogenase molybdate protein
173							nitrogenase molybdate protein
174							nitrogenase molybdate protein
175							nitrogenase molybdate protein
176							nitrogenase molybdate protein
177							nitrogenase molybdate protein
178							nitrogenase molybdate protein
179							nitrogenase molybdate protein
180							nitrogenase molybdate protein
181							nitrogenase molybdate protein
182							nitrogenase molybdate protein
183							nitrogenase molybdate protein
184							nitrogenase molybdate protein
185							nitrogenase molybdate protein
186							nitrogenase molybdate protein
187							nitrogenase molybdate protein
188							nitrogenase molybdate protein
189							nitrogenase molybdate protein
190							nitrogenase molybdate protein
191							nitrogenase molybdate protein
192							nitrogenase molybdate protein
193							nitrogenase molybdate protein
194							nitrogenase molybdate protein
195							nitrogenase molybdate protein
196							nitrogenase molybdate protein
197							nitrogenase molybdate protein
198							nitrogenase molybdate protein
199							

## ALIGNMENTS

## RESULT 1

C;Species: *Caenorhabditis elegans* - *Caenorhabditis elegans*  
 T26930 C;Accession: T12552  
 T26930 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 R;McMurray, A. Submitted to the Protein Sequence Database, June 1999  
 A;Reference number: Z17527  
 A;Accession: T26930  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-684 <WIL>  
 A;Cross-references: UNIPROT:Q45947; EMBL:AL021492; PIDN:CAA16378.1; GSPDB:GN00022; CESP:  
 A;Experimental source: adult uterus; clone Y45F10D  
 C;Genetics:  
 A;Gene: CESP145F10D.3  
 A;Introm: 61/3; 112/1; 142/3; 178/3; 230/3; 289/2; 611/2  
 C;Superfamily: polypeptide N-acetylgalactosaminyltransferase  
 Query Match 42.5%; Score 1392.5; DB 2; Length 684;  
 Best Local Similarity 48.8%; Pred. No. 1.6e-104; Indels 31; Gaps 14;  
 Matches 279; Conservative 87; Mismatches 1; Gaps 14;  
 QY 49 PAAAGGSHSRQQKTFPLGQOKLWDKEDARRDQVRVNGEOPRYPMTDAERVDQ-- 106  
 Db 121 PPAALGDEALDPPEPKYGRH-EKIK-WDEDAVEKEKRRSGPGENGKPKVLPEDKEVEKA 178  
 QY 107 -AYRENGENIYVSDKISLNRSLPDRPHNCNSKYLETLPNTSTIIPPHNEGSSLRT 164  
 Db 179 LSLYKANGTYNAYISDMISLNRSKIDRKKEKNMYSAKLPTVSVIIPPHBEBNLLRS 238  
 QY 165 VHSVNRSPPELKEILVDDPSEKPALRQLEDFLKKNKIDHIVKVLRTKKGRLR 221  
 Db 239 VYSVNRSPPELKEILVDDPSEKPALRQLEDFLKKNKIDHIVKVLRTKKGRLR 298  
 QY 222 MIGASVATGIVTFLSHCEANVNLPLDRIARNRKTIVCPMDVIVDHDFFRYETQAG 281  
 Db 299 QLGQADATGBILIFDAHSENANVNLPLDPIADBDYRTWCPEFDVVIDCTEVRPQ-D 357  
 QY 282 DAMRGAFDWMWVYKIPKIPPIPELQKADPSDFPESPMAGGLFAVDKWFMELGGDPGLEI 341  
 Db 358 EGARGSFDFWAFNPKLPLTKK-DRESPTKFNPSMAGGYFAISAKWFMELGGDVEGLDI 416  
 QY 342 WGGGROYEISPKVWMCGRMEDIPCSRVGHYR-KVYPKV-VPAGSLARNLKVRAEWMD 399  
 Db 417 WGGEYQELSPKFWQHGRGMVDAPCSFRVHRYCKTAPFKWGMGDFVSKRNKRVAEWMD 476  
 QY 400 EYAEVYIQRPEYRHLASGIVAVQVKLRSINCKSFKWMKTMKIAWDLPKYPPYPPA 459  
 Db 477 DYKETLYKHPGVGNADGDKLKGIRECLQKSFDWENKEIAFDQDKYPPAVEPKASA 536  
 QY 460 WGEIRNGVGLCADTKHAGLGSPLRLEGCVR---GRGEAMANNQVFTFTWREDIREGD 515  
 Db 537 EGBIRNGVGFNCIDTFQFKEMQNRGFLRKCTSSDKDGGGE---QDLSLTRWDIRP-- 588  
 QY 516 FOHTKRPCKPFAISHT--SPVTLYDCHSMKCNQFLWYCR-KDTLWMPVSSCMD-SESDR 572  
 Db 589 -KGRKICFDCSISVTDKAPYLFDCHSMKCNQFLKVRQAOKQIYHPIGQDADENGKG 646  
 QY 573 -I FMVTCNPSSLSQTQWLFENTNSTLKEKRN 603  
 Db 647 FLHMKKCDSSSDLQKWAQTDNELLETROAN 678

## RESULT 2

RESULT 2  
 TU2552 hypothetical protein DK2P586H0623.1 - human (fragment)  
 C;Species: Homo sapiens (man)

C;Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 09-Jul-2004

C;Accession: T12552

R;Anborg, W.; Wirkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

Submitted to the Protein Sequence Database, June 1999

A;Reference number: Z17527

A;Accession: T12552

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A;Gene: DKFZP586H0623.1

A;Superfamily: polypeptide N-acetylgalactosaminyltransferase

C;Note: DKFZP586H0623.1

A;Molecule type: mRNA

A;Residues: 1-150-151-276 <ANS>

A;Cross-references: UNIPROT:Q9YH4; EMBL:AL096739

A;Experimental source: adult uterus; clone DKFZP586H0623

A;Note: the cDNA sequence contains a -1 frameshift near codon 150

C;Genetics:

A

F,9-28/Domain: transmembrane #status predicted <TMM>  
 F,29-559/Domain: endoplasmic reticulum luminal #status predicted <LUM>  
 F,51,41,541,559/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F,117,118,288/Binding site: carbohydrate (Thr) (covalent) #status predicted  
 F,119/Binding site: carbohydrate (Ser) (covalent) #status predicted

Query Match 34.3%; Score 1125; DB 2; Length 559;  
 Best Local Similarity 44.0%; Pred. No. 5.6e-83; Mismatches 232; Conservatve 87; Indels 38; Gaps 16; Matches 230; Conservative 87; Mismatches 170; Indels 38; Gaps 16;

Qy 88 GNGEOGRP--YPMTDAAEYDQAVERENGNIYVSPDKISLNRSLPDIRHNCNSKYLETP 145  
 Db 57 GPGEMGKPVVPIKEDQEKMKEMKINQENLMASEMIAINRSLPDIRHNCNSKYLETP 116

Qy 146 NTSLIIPFNEGMSLLRHTVHSNRSPELVAEIVYDPSRERHKKPLEYM-ALFP 145  
 Db 117 TTSVIVVHNEANSTLRLRVHSVNRSPHMISBIVLVDASERDPLKPLSYVKLKV 176

Qy 205 SVRLRTKREGGLTRTRMIGASVATGDTFLDSHCEANVNWAPLDRIAURKTVCP 264  
 Db 177 PPHVIRMEQSRGLIARLKGAAGSKQGQVITFLDAACTGVWLEPLIARIKHDRTVCP 204

Qy 265 MIDVHDHDFRYETOAG-DAMRAFDWEMYKRIPPP--EIQKADSDPFPSPVMAGG 320  
 Db 237 IDVSDTPEY-MAGSDMTYGGFNFNKLNFNRLPVPREMDDRRKGDRKLPLVPTMAGG 294

Qy 321 LFAVDRKFWELGGDPGEIWMGCGQEYELSPKWMGGERMEDIPCSRQHIIKVKVPUKV 380  
 Db 295 LFSIDRDFYFOBIGYDAGMDIWGENLETSFRIWQCGGLEIVTCSHVGHVFRKATPYP 354

Qy 381 PAGVS-LARNLKVVAEWMDEAYIYQRRPBYRLSAGDVAQKKRSSLACKSKFNM 438  
 Db 355 PGEGQIQLINKNRLLAEVWMDERKPNFYIISPGTKVDPGDISRVRGHKLOCKPFSWY 414

Qy 439 MTKLAWD--LPKEYPPVSPAAWGEIRNGTGICADTRKGAGGSPLEQCYGRGSA 496  
 Db 415 LENIYDPSQIPIHY-----FSLGEIRNVEHOCLDNMARKENEKVQIFNC-HGMG-- 463

Qy 497 WNNNOVFTWREDIRPDPQHMKFCDAISHTSPVLYDCISMKGHOLWKYRKDK-TL 555  
 Db 464 --GNQFVSYTANKBIRTD-----LCLDVSKUNGPVMLKCHHLKGHOLWEDPVKULL 515

Qy 556 YHPVSGSCMD-CSBSDHRI-FMTCNPNSLITQQLFEHTNSTLEKF 600  
 Db 516 QHVNNSQCLDKAKTDEDSQVPSIRDCN-GSRSQQLLR-NVTLPEIF 559

## RESULT 4

A45987 polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) - bovine  
 C;Species: Bos primigenius tauris (cattle)  
 C;Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 09-Jul-2004  
 C;Accession: A45587; A48530  
 R;Homa, F.L.; Holland, T.; Lehman, D.J.; Thomsen, D.R.; Elhammer, A.P.  
 J. Biol. Chem., 268, 12609-12616, 1993  
 A;Title: Isolation and expression of a cDNA clone encoding a bovine UDP-GalNAc:polypepti  
 A;Reference number: A45987; MUID:93286099; PMID:7683345  
 A;Accession: A45587  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-559 <HAG>  
 A;Cross-references: UNIPROT:O61397; EMBL:AF031841; NID:93047202; PIDN:AAAC13677.1; PID:93  
 A;Gene: gly-7  
 C;Keywords: glycosyltransferase, hexosyltransferase  
 C;Superfamily: polypeptide N-acetylgalactosaminyltransferase

Query Match 34.0%; Score 1113; DB 2; Length 601;  
 Best Local Similarity 42.0%; Pred. No. 5.8e-82; Mismatches 226; Conservatve 90; Indels 36; Gaps 13; Matches 225; Conservative 90; Mismatches 166; Indels 36; Gaps 13; .

Qy 68 GOKLKDWDKEAIRRDAQGVNGEBOGRGYPMTDAERY--DORYRENGNIYVSPDKISLN 124  
 Db 79 GSEIIGNYSPREPEIPSNGO--PGEHKGKVPVTDDEGMAGRAKEKEFPNTVSDMISNN 135

Qy 125 RSPDPIRPNCSKRYLTLPNWILIPFHNEGWSLLRTVHSNRSPELVAEIVYD 184

C;Keywords: colostrum; glycosyltransferase; hexosyltransferase

Query Match 34.0%; Score 1115; DB 2; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 3.6e-82; Mismatches 230; Conservatve 89; Indels 38; Gaps 16; Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

## RESULT 5

T42251 polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) 7 - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C;Accession: T42251  
 R;Hagen, F.K.; Nehrke, K.  
 J. Biol. Chem., 273, 8268-8277, 1998  
 A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine:Polypept  
 A;Reference number: Z22126; MUID:98192620; PMID:9555933  
 A;Accession: T42251  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-601 <HAG>  
 A;Cross-references: UNIPROT:O61397; EMBL:AF031841; NID:93047202; PIDN:AAAC13677.1; PID:93  
 C;Genetics:  
 A;Gene: gly-7  
 C;Keywords: glycosyltransferase, hexosyltransferase  
 C;Superfamily: polypeptide N-acetylgalactosaminyltransferase

Query Match 34.0%; Score 1113; DB 2; Length 601;  
 Best Local Similarity 42.0%; Pred. No. 5.8e-82; Mismatches 226; Conservatve 90; Indels 36; Gaps 13; Matches 225; Conservative 90; Mismatches 166; Indels 36; Gaps 13; .

Qy 68 GOKLKDWDKEAIRRDAQGVNGEBOGRGYPMTDAERY--DORYRENGNIYVSPDKISLN 124  
 Db 79 GSEIIGNYSPREPEIPSNGO--PGEHKGKVPVTDDEGMAGRAKEKEFPNTVSDMISNN 135

Qy 125 RSPDPIRPNCSKRYLTLPNWILIPFHNEGWSLLRTVHSNRSPELVAEIVYD 184

Db	136 RTIDIRPERRQKHDYPERKLPTVSVVVFHNEGHTPLRTVHSVLRLPPELBBQVWMD	195	Db	373 FDWGGGENELSKIMMGGTURIPCGSHVGHFRKSPYKRTGVMVLRKNSIRLAVW	432
Qy	185 DFSREHHLKPLDYMALF-PSVRLRTKRGRLRT	243	Qy	398 MDEAYEYTYYORRPEYRHLSSAGDVAVQKUSSLRCKSNTKFWMTKIAWDLPLKPPVVERPA	457
Db	196 DSDSKPHLKEKLDKVKTRKGKVTWTRTEQREGLINARSIGAKHSTGEVVLFLDAHCEVN	255	Db	433 LDDKYTYTVERINNQLG-DFGDISSRKLRBEGACKSKFWYLDNI--YPELFVPG-E-S	486
Qy	244 VNMPLPLDRARAKRTIVCPMDVHDPRYRTOAGD--MKGADDMWYKRIPI	300	Qy	458 AANGETRIVGTLGICADTHGALGSPRLREG---CYGRGRGEAUNNNQVFTFTWBDIR	512
Db	256 TNWLPLPLPPLAKRKRKVNUTPVVIDGIDSNSWEYRSVYGSNPNAHSGIFEWGLLYKETOT	315	Db	487 VAKGBLRNQTSOCLDS--AVGREVENKAITPYCPHCBQGNGOYW---MUSKDGIR	537
Qy	301 PE-LQKADPSDPESPVMAGGLFAVDRKMFELLAGLPGLEIMWGQEBSFRKWMCG	358	Qy	513 PGDQHTRKCFDADSKHSPVTLYDCHSMKGNOIWKDK-TLYHPSGSCMDSSDH	571
Db	316 EREFAHARKNSQPRSPHTAGLFAINRNLWKEGYDGGIQLINGEQEYLSFRKWMCG	375	Db	538 RDSS----CVYAG--SDVWTFPCBOMKGNGQEWVYHDTGRLQHAVSQKCLGTMKOGA	589
Qy	359 RMEDIPCSRVGHTYRKVYVY--KYPAGVSLARNLKVRAEVNMDYEAETYORRPEYRL	415	Qy	572 RIFKNTCPSSLTOQWLAEBTN	593
Db	376 GIVFVPCSHVGHVYRSHMPVSPGKPSGKPEVISMNMRVVKTMWDYSKYLTREPOATV	435	Db	590 KLEMVACQYDDPFOHMKKEIN	611
Qy	416 SAGDVAQVKLRSKSLNCKSKFKWPMKIKLADLPKPYPPVPAWGEIRNVCGLADPK	475			
Db	436 NPGDISAQLAIRDKLOCKSFKWKYMEVNAVTDVLSKPYML-RPDNYWGEARNPATGKCLD-R	493			
Qy	476 HGAIGSPLRLEGCVGRGRGEAAMNNQVFTFTWREDIREPDPQHFKCDAISHTSPVTL	535			
Db	494 MGGIPGPMGATGC-----HGGGNQLIRLNQVQOMAQBR-----WCITA---NGIRI	537			
Qy	536 YDCHSMKG--NQLNKY-RKDQTLYTHPVSSGSCMDSCESDRHIFMNTCNPSLTOQWLF	590			
Db	538 QANHCVKGTVNGFWPSYDRKTKQIHSQKRCITVSESGSEVTLQCTEDBMRQKFWK	595			
<b>RESULT 6</b>					
22247 polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) - <i>Caenorhabditis elegans</i>					
C;Species:	Caenorhabditis elegans		C;Species:	Caenorhabditis elegans	
C;Accession:	T42247		C;Accession:	T42246	
A;Title:	cdNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept		A;Title:	cdNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept	
A;Reference number:	03-Dec-1999 #text_change 09-Jul-2004		A;Reference number:	03-Dec-1999 #text_change 09-Jul-2004	
A;Molecule type:	RNA		A;Molecule type:	RNA	
R;Hagen, F.K.; Nehirke, K.			R;Hagen, F.K.; Nehirke, K.		
J. Biol. Chem. 273, 8248-8277, 1998			J. Biol. Chem. 273, 8248-8277, 1998		
A;Status: preliminary; translated from GB/EMBL/DDBJ			A;Status: preliminary; translated from GB/EMBL/DDBJ		
A;Molecule type: RNA			A;Molecule type: RNA		
A;Residues: 1-624 <HAG>			A;Residues: 1-626 <HAG>		
A;Cross-references: UNIPROT:061392; EMBL:AF031836; NID:93047192; PIDN: AAC13673.1; PID:93			A;Cross-references: UNIPROT:061392; EMBL:AF031836; NID:93047192; PIDN: AAC13672.1; PID:93		
C;Genetics:			C;Genetics:		
A;Gene: GLY-5			A;Gene: GLY-5		
C;Superfamily: polypeptide N-acetylgalactosaminyltransferase; hexosyltransferase			C;Superfamily: polypeptide N-acetylgalactosaminyltransferase		
C;Keywords: glycosyltransferase; hexosyltransferase			C;Keywords: glycosyltransferase; hexosyltransferase		
Query Match 32.9%; Score 179; DB 2; Length 626;					
Qy	Best Local Similarity 44.9%; Pred. No. 3.5e-79;		Qy	Best Local Similarity 44.9%; Pred. No. 3.5e-79;	
Db	Matches 227; Conservative 75; Mismatches 161; Indels 40; Gaps 14;		Db	Matches 227; Conservative 75; Mismatches 161; Indels 40; Gaps 14;	
Qy	103 RVDQAYRENGFNTIVSKLISLNSLPLDPRHPCNSKRYLETPLNTSIIIPFNGWSSL	162	Qy	103 RVDQAYRENGFNTIVSKLISLNSLPLDPRHPCNSKRYLETPLNTSIIIPFNGWSSL	162
Db	133 KYDKGMNNAHFNOYASDMDVHRTLPTNTDAECKTEKVNENLPTSVICFHNEAWSVLL	192	Db	133 KYDKGMNNAHFNOYASDMDVHRTLPTNTDAECKTEKVNENLPTSVICFHNEAWSVLL	192
Qy	163 RTVHSVLNTSPPPEVAVTIVLVDSDRSRBLKKPLDYMALF-PSVRLRTKRGRLRT	221	Qy	163 RTVHSVLNTSPPPEVAVTIVLVDSDRSRBLKKPLDYMALF-PSVRLRTKRGRLRT	221
Db	193 RTVHSVLNTSPPPEVAVTIVLVDSDRSRBLKKPLDYMALF-PSVRLRTKRGRLRT	252	Db	193 RTVHSVLNTSPPPEVAVTIVLVDSDRSRBLKKPLDYMALF-PSVRLRTKRGRLRT	252
Qy	222 MLGASVATDVITLSDHCEANVNLPLPLDRARAKRTIVCPMDVHDPRY-ETOA	280	Qy	222 MLGASVATDVITLSDHCEANVNLPLPLDRARAKRTIVCPMDVHDPRY-ETOA	280
Db	253 LRGAVATRVLVLTSDHCECMEGWMEPLDRKRDPTTVCPVIDVNDNTPEYHSSKA	312	Db	253 LRGAVATRVLVLTSDHCECMEGWMEPLDRKRDPTTVCPVIDVNDNTPEYHSSKA	312
Qy	281 GDAARGAFDMEWYKRIPIPPELOK--ADPSDPESPVMAGGLFAVDRKMFELLAGLPG	338	Qy	281 GDAARGAFDMEWYKRIPIPPELOK--ADPSDPESPVMAGGLFAVDRKMFELLAGLPG	338
Db	313 YFTSVGGFDGQLOFNWHSPERDRKRTTRPDRPVSPTWAGGLPSIDKEYFEKLGTYPG	372	Db	313 YFTSVGGFDGQLOFNWHSPERDRKRTTRPDRPVSPTWAGGLPSIDKEYFEKLGTYPG	372
Qy	339 LEIINGGQEVSIFSKVWMCGRMEDIPCSRVGHTYRKVYKPVAGVS-LARNLKVRAEW	397	Qy	339 LEIINGGQEVSIFSKVWMCGRMEDIPCSRVGHTYRKVYKPVAGVS-LARNLKVRAEW	397
Db	373 FDINGGENELSKIMMGGTURIPCGSHVGHFRKSPYKRTGVMVLRKNSIRLAVW	432	Db	373 FDINGGENELSKIMMGGTURIPCGSHVGHFRKSPYKRTGVMVLRKNSIRLAVW	432
Qy	458 AAWGBIRWVGTGLGACTKRGALGSPRLRGCVGRGRGAUNNNQVFTF-----WR	508	Qy	458 AAWGBIRWVGTGLGACTKRGALGSPRLRGCVGRGRGAUNNNQVFTF-----WR	508
Db	487 VAKGBVRNSAV-----OPARCLDMVGRHE--KNRPGVTQCHGOGGNQWM	531	Db	487 VAKGBVRNSAV-----OPARCLDMVGRHE--KNRPGVTQCHGOGGNQWM	531
Qy	509 EDIRGDPHKKCFDADSHTSPTVLDLCHSMKGNOIWKDK-TLYHPSGSCMDSSDH	567	Qy	509 EDIRGDPHKKCFDADSHTSPTVLDLCHSMKGNOIWKDK-TLYHPSGSCMDSSDH	567







Qy	97	DW-----GEAGGAGVSHLIPQEOKLADSTAVNQFNLVPGISVRSLPFR	Db	377	PKVPGAVS-LARNLKRYAEVNNDYEYIYORRPE-----YHLSACDVAVOKKLSSL
Qy	132	HPIGNSKRYLETIPNTSIIPFNNGWMSLRLTRHVSUNRSPPELVATIVLVDPSDREH	Db	422	PHTPKGIVQIANQVRLAEVNNDYEYIYORRPE-----YHLSACDVAVOKKLSSL
Qy	144	KPSCKRNNTYPDNLNPTTSVII-----	Db	422	PHTPKGIVQIANQVRLAEVNNDYEYIYORRPE-----YHLSACDVAVOKKLSSL
Qy	192	LKRP-LEDWALPFS-VRLRTRKREGRLTRMLGASATGDIVTFLDSHCEANHGWLP	Db	432	CISPKWFMKTIKAWDLPKRPPPEPAWGERANVGLCALT-KHAGLSURLEGCVR
Qy	204	LRYPTLDTKPKPTDIKISKERRGLIARANGQAQAGDVLTFLASHCETKGWLP	Db	492	CQFTWVNTL--YPAEXVPLNPVTS-GYKSVGSPCLVUGENNOGKGLILYTC-H
Qy	250	LUDRIRARRKTYCIPMIVDHDFFRYETOAGDMGARDWEMYKRIPIPPELOK--A	Db	491	GRBAAMNNMQVTFWREDIPGDPQIT--KXCFDAISHTSPV-TLYDERS--MKGK
Qy	264	LLTRIKLNKAVKPCVVIDINDNFOYO-QGIEWFRGGENWNLQFRWYGMPTAMAKHIL	Db	537	GLG----GNOFEYSAOBR----HNQKBLCLHATQGVQLKACVYKQHRTIARBE
Qy	307	DPSDPFESVUMAGLFLAIDRKWFLWELLGDPGLBWIWGGQEVESPKVMMGGMEDI	Db	545	QIWKYKOKTLPHVSCCMDSESDTRIFMTCNPESLTLQWLP
Qy	323	DPTGPPIESPPTMAGLFSINRNVEPELGEVDPGMDIWGENLESPRIMOCGSEV	Db	587	QINWIRKDQNLNPFLKMCLS-SNGEHPNLV-PCDATDLOLQWLP
Db	443	RVELRKKLNCKSFKWNLYQVQD	465		
RESULT 15					
Qy	367	RVGHTYRKVPUYKPP--AGVSLARNKKEAVEMDEAVYIYORRPE-YHNSAGDAV	Db	383	HVGHVFVFKSSPHDFGGKSSGVNLNLLRAEVWMDWKGHVFYKLAPOAHRSKSIDVE
Qy	423	OKKURSSANCKSKPWFEMTKLAWD	445		
Db	443	RVELRKKLNCKSFKWNLYQVQD	465		
RESULT 16					
Qy	1549	polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) 9 - Caenorhabditis elegans	Db	443	polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) 9 - Caenorhabditis elegans
Qy	1549	NALternate names: protein-UDP acetylgalactosaminyltransferase	Db	443	NALternate names: protein-UDP acetylgalactosaminyltransferase
Qy	1549	Species: Mus musculus (house mouse)	Db	443	Species: Caenorhabditis elegans
Qy	1549	Date: 13-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004	Db	443	Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
Qy	1549	CAccession: JG5247	Db	443	CAccession: T31549, T42253
Qy	1549	R.Zara, J.; Hagen, F.K.; Ten Hagen, K.G.; Van Wuyckhuyse, B.C.; Tabak, L.A.	Db	443	R.Matthews, L.
Qy	1549	Biochem. Biophys. Res. Commun. 228, 38-44, 1996	Db	443	submitted to the EMBL Data Library, October 1999
Qy	1549	A.Title: Cloning and expression of mouse UDP-GalNAc:polypeptide N-acetylgalactosaminyltr	Db	443	A.Reference number: T31549
Qy	1549	A.Reference number: JG5247; MUID:97069650; PMID:9912633	Db	443	A.Accession: JG5247
Qy	1549	A.Molecule type: mRNA	Db	443	A.Molecule type: DNA
Qy	1549	A.Residues: 1-633 <ZAR>	Db	443	A.Residues: 1-579 <WIL>
Qy	1549	A.Cross-references: UNIPROT:P70419, GB:U70538; NID:915757; PIDN:AA09579.1; PID:915757	Db	443	A.Cross-references: UNIPROT:Q9u2c4; EMBL:AL117202; PIDN: CAB57897.1; CESP:Y47D3A.23
Qy	1549	A.Experimental source: testis	Db	443	A.Experimental source: clone Y47D3A
Qy	1549	C.Comment: This enzyme catalyzes the formation of N-acetyl-D-galactosamine alpha-O-Ser/T	Db	443	R.Hagen, F.K.; Nehrke, K.
Qy	1549	Initiating O-glycosylation of serine and threonine residues on an array of glycoproteins.	Db	443	J. Biol. Chem. 273, 8268-8277, 1998
Qy	1549	C.Superfamily: polypeptide N-acetylgalactosaminyltransferase	Db	443	A.Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine:Polype
Qy	1549	C.Keywords: glycosyltransferase; hexosyltransferase	Db	443	A.Reference number: Z22126; MUID:98192620; PMID:9525933
Qy	1549	Query Match 28.5%: Score 934; DB 2; Length 633;	Db	443	A.Accession: T42253
Qy	1549	Best Local Similarity 39.2%; Pred. No. 2e-67; Matches 179; Indels 42; Gaps 18;	Db	443	A.Status: preliminary; translated from GB/EMBL/DDBJ
Qy	1549	Matches 206; Conservative 98; Mismatches 19; Indels 42; Gaps 18;	Db	443	A.Molecule type: mRNA
Qy	1549	90 GEOGRPYPTDAERVQDQVRENG----FNIYVSDKISIANSU-PDIRHPCNCSKRY-L	Db	443	A.Residues: 1-202, D, 204-579 <HAG>
Qy	1549	122 GASCKPKITIHLSSRPEQKEKERGETKHCNAPASDRDRSHDGPDTREPECEQKFRK 181	Db	443	A.Cross-references: EMBL:AF031843; NID:93047206; PIDN: AAC13679.1; PID:93047207
Qy	1549	142 ETLNTSIIIPFRNGBWSSLRLTRHVSUNRSPPELVAEVLUDDPSDRSHKKELEDYNA 201	Db	443	C.Genetics:
Qy	1549	182 PPLFTSVIIVFNAEWSTLRTVHSVLYSSPAILIKEILVDDAHSVDDLHEKLEEVK 241	Db	443	A.Gen: CESP:Y47D3A.23; gIV-9
Qy	1549	202 LFPSPRILATKPKRKGSLRTRMLGASVATDVITDSDHEANVWLPPLDRFARNKTI 261	Db	443	A.Introns: 45/1; 179/1; 276/1; 343/3; 373/1; 453/3; 491/3; 533/3
Qy	1549	242 QFSIVKIVQROERKQKLITLGLAAVATBTLTFLDAHCFCYGSWPLIARTANNTAV 301	Db	443	C.Superfamily: polypeptide N-acetylgalactosaminyltransferase
Qy	1549	262 VCPMPMDVHDDFRYETQA--GDMRGADPWEWY--KRIPPFELQKADPSPFESPV 316	Db	443	C.Keywords: glycosyltransferase; hexosyltransferase
Qy	1549	302 VSPDPIASIDANTPFENKPKPYGSMNHRGFDWLSFGWSLDPDKRKEKDTEPKIPT 361	Db	443	Query Match 27.6%: Score 906; DB 2; Length 579;
Qy	1549	Best Local Similarity 40.2%; Pred. No. 3.3e-65; Matches 212; Conservative 88; Mismatches 182; Indels 46; Gaps 22;	Db	443	Best Local Similarity 40.2%; Pred. No. 3.3e-65; Matches 212; Conservative 88; Mismatches 182; Indels 46; Gaps 22;
Qy	1549	94 QRVNGEGRPYPT--DAERVQDQVRENG--FNIYVSDKISIANSU-PDIRHPCNCSKRY 140	Db	443	Query Match 27.6%: Score 906; DB 2; Length 579;
Qy	1549	141 L-LETPNTSIIIPFRNGBWSSLRLTRHVSUNRSPPELVAEVLUDDPSDRSHKKELEDYNA 140	Db	443	Best Local Similarity 40.2%; Pred. No. 3.3e-65; Matches 212; Conservative 88; Mismatches 182; Indels 46; Gaps 22;
Qy	1549	129 DYALPKTISVIFTDEANTPLIARTVHSVINSRSPPELQEVILDDNSKSQLQPLDHE 188	Db	443	Query Match 27.6%: Score 906; DB 2; Length 579;
Qy	1549	200 MALP-PSYRILTRKREGRLTRMLGASVATDVITDSDHEANVWLPPLDRFARNR 258	Db	443	Best Local Similarity 40.2%; Pred. No. 3.3e-65; Matches 212; Conservative 88; Mismatches 182; Indels 46; Gaps 22;
Qy	1549	189 KARGGKVKLIRKTRHVSUAKLAGAREBAGDILVFLDSHCEANHGWLPPLDRFARNR 248	Db	443	Query Match 27.6%: Score 906; DB 2; Length 579;
Qy	1549	259 KTIYCPMIDVHDDFRYETQA--AMRGADEWY--KRIPPFELQKADPSPFESPV 314	Db	443	Best Local Similarity 40.2%; Pred. No. 3.3e-65; Matches 212; Conservative 88; Mismatches 182; Indels 46; Gaps 22;
Qy	1549	249 TAIVCPMIDSIDNTLAVH--GDWSLSTGGFWALHFTWEGSLSBECOKRKEPDYRS 305	Db	443	Query Match 27.6%: Score 906; DB 2; Length 579;
Qy	1549	315 PVWAGGLEFVDRKWFNWLGGDGPGLWGGCQEYTSFKVWCGGRMHDIPCSRVGHMYRK 374	Db	443	Best Local Similarity 40.2%; Pred. No. 3.3e-65; Matches 212; Conservative 88; Mismatches 182; Indels 46; Gaps 22;
Qy	1549	306 PTMAGGLLAAANREYFFVGGYBSEMDINGGENLBISFRAMCGGSIEFTPCSHVGHFRA 365	Db	443	Query Match 27.6%: Score 906; DB 2; Length 579;
Qy	1549	375 YVPKVPGAVS---LARNLKRYAEVNNDYEYIYORRPEIRHLSAGDVAVQKLRSL 430	Db	443	Best Local Similarity 40.2%; Pred. No. 3.3e-65; Matches 212; Conservative 88; Mismatches 182; Indels 46; Gaps 22;

RESULT 17

T42244 probable polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) - *Caenorhabditis elegans*  
 C;Species: *Caenorhabditis elegans*  
 C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C;Accession: T42244  
 R;Hagen, F.K.; Nehrke, K.  
 J. Biol. Chem. 273, 8268-8277, 1998  
 A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine:Polypept  
 A;Reference number: 222126; MUID:98192620; PMID:9525933  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Residues: 1-589 <HAG>  
 A;Cross-references: UNIPROT:061390; EMBL:AF031834; NID:93047188; PIDN: AAC13670.1; PID:93  
 C;Genetics:  
 C;Gene: GLY-4  
 C;Superfamily: polypeptide N-acetylgalactosaminyltransferase  
 C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 25.9% Score 883; DB 2; Length 589;  
 Best Local Similarity 37.4%; Pred. No. 2.5e-63; Mismatches 80; Indels 82; Gaps 18;  
 Matches 209; Conservative 80; Mismatches 188; Indels 82; Gaps 18;

Qy 56 HSRQK-----KTR-----FLGDGOKLKDWDKEAIRRDAQVRGVNGBQGRPPYPTDAERVDQ 106  
 Db 83 HERTEKDTWKTIDVEKELNSK-----WH-----ED-----OG-----ED 111

Qy 107 AYRENGNYIVSISKISLNSLSPDPRHPNCNSKEYLET-LPNNTIIIPFNEGWSLLRTV 165  
 Db 112 KYQNSFNQBAQDNLNPTKRPDSREPOCDVDSKVGRQPTVITYHNEASSLLRTV 171

Qy 166 HSULARSPELVAIIVDPSRERHKKPLDYMALFPSVRLRTKREGIRTRMG 225  
 Db 172 FSYVNQSPBELLBLBIVLVDDNSDQVETKEL---AQQTIRVTLRNNMREGIJSRRTGKA 227

Qy 226 SVATGDVITFLDSHCEANNNWLPLDRTARRNRKTIQPMIVDIDHDDFRTYETQAGAMR 285  
 Db 228 QVARAPVUTELDHSIECKWMLPRLAIAENKAVANAPIVINDVNFYAGAS-LR 286

Qy 286 GAFDWEMYKTRIPPELQK---ADPSDPFSPVMASSFLAVDPRKWFELGSDPGLRW 342  
 Db 287 GSFDTWTLVFRWEPNNEQKERHKAHTAPIRSPTMAGFLFASKEWNEBLGTYIDLMRW 346

Qy 343 GEGOYELSPKVMKGGRMEDIPPSRVRGHYKRVYKPGAS---LARNLKVAEVNDE 400  
 Db 347 GEGENLEMSFRVWQCGGSLEIMPCCSRVGHYKRVYKPGAS---LARNLKVAEVNDE 406

Qy 401 YAFYIYQRPYRPHLSAGDVAVQKLRSSLNCKSPKWMKIAWDLPKYPVPEPPAAW 460  
 Db 407 YAKIYLKQVPSAFTVNFEDITDILARBLQCLSPKWLNT---YPOLEIPKTP 459

Qy 487 GYVGRGRGERAAWNMNOVFTPTWREDIRPQDQHTKFCFDALSHTSPTVLYDCHSMKG-N 544  
 Db 482 HC-QSGK---SSPQMLSUSKEGNLR---RENTC---ASERGNIRMKTC-SKKAQFN 527

Qy 545 QWVYKQKTKLWVPSGSCMDCE---SDHTRPNTCPSSLTQOWLF 589  
 Db 528 ERWAY-ENKMRNLKSGKCMSTANLKGDNATVWE-CDEKDSHOKNP 573

RESULT 18

T27397 hypothetical protein Y75B8A.9 - *Caenorhabditis elegans*  
 C;Species: *Caenorhabditis elegans*  
 C;Accession: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 A;Cross-references: UNIPROT:Q9XW72; EMBL:AU033514; PIDN:CAA22098.1; CESP:Y75B8A.9  
 C;Genetics:  
 A;Gene: CESP:Y75B8A.9  
 A;Accession: T27397  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-605 <WIL>  
 A;Experimental source: clone Y75B8A  
 A;Reference number: 220361  
 A;Intromb: 78/3; 251/3; 370/3; 391/2; 529/3; 565/3  
 C;Superfamily: polypeptide N-acetylgalactosaminyltransferase

Query Match 24.4% Score 801; DB 2; Length 605;  
 Best Local Similarity 31.9%; Pred. No. 1.1e-66; Mismatches 102; Indels 103; Gaps 18;  
 Matches 187; Conservative 102; Mismatches 194; Indels 103; Gaps 18;

Qy 80 IRRDAQRYNGNGQR-----QYRENGFNTIVSDKISLNSLDPTRHPNCNSKRYVLTPTSI 149  
 Db 104 DLGKINGKAEDDLQVQEGKYKQYQFNGLSDRGSRKIKDSRNARCSLTYSDSLPASI 163

Qy 150 IIPFHNEGWSLLRTVHSVNLNSPPELVAETVWVDDSD---REHUKPLADYMAEPS 205  
 Db 44 IOKHEKLUNGQGQGRHDEDDDEGAKEDEDAVEKQNTAAPPPLKSFTFPDRSKETBDT 103

Qy 166 -----QYRENGFNTIVSDKISLNSLDPTRHPNCNSKRYVLTPTSI 149  
 Db 164 WCYFNNSPSVLRMVISJFDTCRHEHILVDSSEWSNATDEAKYKRIHIL 223

Qy 206 VRLTTRKREGIITRTRMAGSASTGTVTFLSHCEANNNWLPLDRTIARNRKTIYCPM 265  
 Db 224 VFKLTDOKNEGLTRAKLFGARRANGEVILVFLDSHCEVNBEWLPLDQIKQSRVYVCP 283

Qy 266 IVDIDHDDPRTYQAGAMRGARDMENYKTRIPPLQKADS---DPFESPVAGOLF 322  
 Db 284 IDTIDATMVKY---VESPVCTGCVNWAMTPK-WDYPHRYSYFEDPMVNVPLKSPPTMAGLF 340

Qy 323 AVDRKWFELGGDGPGLIWIWGGDQYETSPKVMCGGRMEDIPCSRVQHLYKRVYKPGKA 382  
 Db 341 AIDKBYFEEIGSYDGEHDWVGAENVE-SVRIWTCGGLIMCSRVHAFRQRPSIKT 400

Qy 383 GVSLLARLKRVABWMMBDEAYTIIYQRFPEYRILS-ACDVAVOKKLSLNLCKSPKWMK 441  
 Db 401 D-SMGKNSVRLARWVLDYLENPFEARPNYRIFTDQGDLTSRISRLRNLOCKPFWIEN 459

Qy 442 IAWDLPKYPVPEPPAAWGERVN---GIG-LCAUTKHG---ALSPSLR 484  
 Db 460 1---YPELUDPNTP---NOLANOLIVAGKYLKIKUNGTHCLSAANSQRIANGNVE 511

Qy 485 LEC-CVGRGERAAWNMNOVFTPTWREDIRPQDQHTKFCFDALSHTSPTVLYDCHSMKG-N 544  
 Db 512 MRKONH---MERMKQWKSSTNEI---PMGSSMCLSLRGISVIL---CHNGAH 559

Qy 545 QWVYKQKTKLWVPSGSCMDCE---SDHTRPNTCPSSLTQOWLF 589  
 Db 560 QWQVSNAGKLY---SRSVNKATGSNDVSAALSTLKFCSLANSFQF 602

T2252  
polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) 8 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T42252; T2290  
R;Hagen, P.K.; Neirke, K.  
J. Biol. Chem. 273, 8248-8277, 1998  
A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine:Polypept  
A;Residues: 1-421 <HAG>  
A;Cross-references: UNIPROT:045293; EMBL:AF031842; NID:93047204; PIDN: AAC13678.1; PID:93  
R;Matthews, L.  
submitted to the EMBL Data Library, October 1998  
A;Reference number: 220338  
A;Accession: T2290  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-421 <WIL>  
A;Cross-references: EMBL:AL032622; PIDN:CAA21500.1; GSPDB:GN00021; CESP:Y66A7A.6  
A;Experimental source: clone Y66A7A  
C;Genetics:  
A;Gene: gly-8; Y66A7A.6  
A;Map position: 3; 62/3; 97/2; 181/3; 226/2; 279/2; 318/3; 369/3  
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 19.3%; Score 632; DB 2; Length 421;  
Best local Similarity 34.0%; Pred. No. 3-46-43; Mismatches 146; Indels 44; Gaps 10;  
Matches 138; Conservative 78; MisMatches 146; Indels 44; Gaps 10;

QY 83 DAQRYVNGSGRGRPY-----PMTDAERVDAYREN-----GENTY 116  
Db 21 BAEKLKCBEDVDPYENLEGMDLKLITE-RKCNHTLKENITTEASRSKKSENGTKSPAFDAL 79  
QY 117 VSDKTSLANSLSDPDRHNPNCNSKRVLETLPTNLIIPFHNEGSSLRTHSVLNSRPPA 176  
Db 80 SEKLGKPNRNGVKOKHKLGEKEEY-DASYSTSUVVHNEALSTLIRMINGTIEPTPSL 138  
QY 177 VAEIVLWDDFSDREH-LKCPLEDY--MALLFSPSVERLTKRREGLIRTRMLGASVATGIV 232  
Db 139 LKEIVLYEDPASEBDVLTKELEFAKIKGLEDKLTIRKBYRQSLIRAKWHSRATGEV 198  
Db 233 TFLDHSCEANVNWLPLDRIARNRKTYCVPMDIVDIDDDPFRYETQADGADMRAFDWEM 292  
Db 199 IVEFDHSCEAVERWIEPFLQPIKEPKSTIUPVUDLNPVSPDFDPSM--VAKSGFDNGF 256  
QY 293 YKKRIPRPLQLKADPSD--PFESPVMAGOLFADVRKWMFELGGYDPGLIWIWGEQEVTS 350  
Db 257 TFKWLYLPWHEFPTPENNVPFNPSPAMPSSGLAMRKEYFVELGEYDGMHEWSENTELS 316  
QY 351 FKVWNGGRNEDIPGSRVGHYIYRKVWVPKYVPAFVGSLA-RNLUKRAEVWMDYEAQYIQR 409  
Db 317 LKAMICGGGRWVACPSRSGHVFMRKRPYTSKPGMDTALIVAVRVAKTWLGEGYESKFPVK 376  
QY 410 PEYRHLASDPAVOKKLRSIINCKSFKWFNTKIANDLPKYPPYER 455  
Db 377 PFGAKNVFGDILTEPMQVKDRIKCKOMKWFENV-----YPELIP 415

RESULT 20  
H97186 glycosyltransferase domain containing protein [imported] - *Clostridium acetobutylicum*  
C;Species: *Clostridium acetobutylicum*  
C;Accession: H97186  
R;Nolting, J.; Breton, G.; Omelchanko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.; Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clo*  
A;Reference number: A96900; MUID:21359325; PMID:21359325

A;accession: H97186  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1044 <KUR>  
A;Cross-references: UNIPROT:097399; GB:AE01437; PIDN: AAK80283.1; PID:915025335; GSPDB:GN  
A;Experimental source: *Clostridium acetobutylicum* ATCC824  
C;Genetics:  
A;Gene: CAC2327  
Query Match 4.5%; Score 147; DB 2; Length 1044;  
Best Local Similarity 20.6%; Pred. No. 0.0025; Mismatches 108; Indels 22; Gaps 8;  
Matches 48; Conservative 55; MisMatches 108; Indels 22; Gaps 8;

QY 141 LETLPNTSIIIPFHNEGSSLRTHSVLNSRPPELVARIIVLDFDS--DREHLKKELB 197  
Db 1 MNNSPKVPSITIWNN-GQHHLKNCFSKSEHKLNS2PSDKIKEYIVUDNGSGSVFELKKN-- 57  
QY 198 DYMALFPPSVERLTKRREGLIRTRMLGASVATGSDVITLDSHEANVNWLPLDRIAR- 256  
Db 58 -----YPAVIIKRDNSGAKNDAAKIAESEYELAIINNDMQLDKWLNDFETLNC 112  
QY 257 NRKTIIVCPMDIVDIDHDDPFRYETQADGADMRAFOWEMYKRPITPPELQKADSPDPFSSPV 316  
Db 113 NDNSYVUCAGSKIVWMDGSKLDFAGGSVSPAGYQVQYDGM---DIBGANKKXNEDDI 167  
Db 168 LFAGGGSMJRKDVFIETGGFDKDYFAY-YEDVDLGWHLWLGYKVR--CSK 217

RESULT 21  
D87531 glycosyl transferase family protein CC2277 [imported] - *Caulobacter crescentus*  
C;Species: *Caulobacter crescentus*  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C;Accession: D87531  
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.; Bambury, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.; Smulders, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.; Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A;Reference number: A87249; MUID:21173698; PMID:11259647  
A;Accession: D87531  
A;Cross-references: UNIPROT:Q9A618; GB:AE005673; NID:913423792; PIDN: AAC24248.1; GSPDB:GN  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-306 <STO>  
C;Genetics:  
A;Gene: CC2277  
C;Superfamily: Rhizobium succinoglycan biosynthesis glycosyltransferase  
Query Match 4.2%; Score 139; DB 2; Length 306;  
Best Local Similarity 25.4%; Pred. No. 0.0019; Mismatches 91; Indels 40; Gaps 10;  
Matches 57; Conservative 36; MisMatches 91; Indels 40; Gaps 10;

QY 144 LPNTSIIIPFH--EGWSSLRTHSVLNSRPPELVARIIVLDFDSREHLKPLEDVA 201  
Db 1 MARVYVUMIPTTORPDPGLAVAARSVYFGQVGDFAEL--SLVIVD--NDVPSAKRVDALR 56  
QY 202 L-FPSVRLTTRKRBGLIIRTRMLGASVATGVDVITLDSHEANVNWLPLDRIARNEK 260  
Db 57 KGACCPVIVNEKEPGVAFARNAGMARAASGDPFAFLDDEAISGWLALLAQERYAD 116  
QY 261 IVCPMI-----DVIDHDF--RYETOQADAMGAFDN----EMYKRPITPPELQKAD 307  
Db 117 WVFGPKARAPAHIDQHDLRERFSRGPQAQVDFIDHYGCGDSLIRSALD----D 170  
QY 308 PSDP-----ESPMAGOLFADVRKWMFELGGYDPGLEW 342  
Db 171 PVAFAVERNFIGGDDLLFGHMGAGKRFWAE----PAAWW 209

RESULT 22

H87306 glycosyl transferase family protein CC0465 [imported] - *Caulobacter crescentus*  
 C;Species: *Caulobacter crescentus*  
 C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
 C;Accession: H87306  
 R;Niemer, W.C.; Peldbryum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
 n, J.; Brzozova, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Ventter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A;Title: Complete Genome Sequence of *Caulobacter crescentus*.  
 A;Reference number: A87249; MUID:21173698; PMID:11259647  
 A;Accession: H87306  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-308 <STO>  
 A;Cross-references: UNIPROT:Q9AX7; GB:AE005673; NID:913421638; PIDN:AAK2452.1; GSPDB:G  
 C;Genetics:  
 A;Gene: CC0465

Query Match 4.2%; Score 128.5; DB 2; Length 308;  
 Best Local Similarity 23.1%; Pred. No. 0.0021; Mismatches 124; Indels 77; Gaps 16;  
 Matches 76; Conservative 52; Mismatches 124; Indels 77; Gaps 16;

```

QY 145 PNTSIIIPPHNEGSSSLRRTVHSVNRSPPEVIAEVVJUDPS--DREHLKPLEDMA 201
Db 20 PNVSVVMMVYRTG-EALVESIRHVL--AEPVDEFIIVDNGSSRDEMLRS----LA 70
QY 202 LP-PSVRLRKTKEKREGILRTMIGASVATGDNVITFLD----SHCEANV----NWLP 248
Db 71 LTPRPRVVKQGKGVNGFARGANIGAVTGGYVIFLNPDANLQPSCAVSLVTAFKGQVLP 130
QY 249 PLIDRIRN-----RKTIVCPMDVIDHDFF--RYETVQAGDAMRGAFDNEYMYRI 297
Db 131 TIVGARVLTGSEQRGGRRGDTPISTVLISGOLTRRYPKLAG----FEIHRNE 182
QY 298 PIPPELOKADPSDPPFESPMPMAGGLFAVURKWFELGGYDPG--LEIWMGEQVQISFKWM 355
Db 183 PL-----PGAVPVPMPTRISGACFAMRRADFAVNGFDEGYFLHV--EDIDLCWRARR 231
QY 356 CGGRMEDIPCSRQHIVYRKVV--PYKVP--AGVSLARNLKVRAEVWMDBEYAVIYQRRE 411
Db 232 AGQVLFQPNAAVYVHLGNTSLELPVKEFHKQGLTR-----YFIRBAD 275
QY 412 YRHLISAGDVAVQKQKRSLLNCKSFKFWFMT 440
Db 276 SLOLFAAVVLLAPAIMLMSVCRPLWKL 304

```

RESULT 23

B84263 succinoglycan biosynthesis protein [imported] - *Halobacterium* sp. NRC-1  
 C;Species: *Halobacterium* sp. NRC-1  
 C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
 C;Accession: B84263  
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.  
 Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabol  
 Jung, K.H.; Alam, M.; Freitas, T. 2000  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181,  
 A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehrhardt, H.; Lowe, T.M.; Li  
 A;Title: Genome sequence of *Halobacterium* species NRC-1.  
 A;Reference number: A84160; MUID:20504483; PMID:11016950  
 A;Accession: B84263  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-328 <STO>  
 A;Cross-references: UNIPROT:Q9HQP3; GB:AB004437; NID:910580616; PIDN:AA19470.1; GSPDB:G  
 C;Genetics:  
 A;Gene: exom

Query Match 4.1%; Score 126; DB 2; Length 328;  
 Best Local Similarity 20.4%; Pred. No. 0.0037; Mismatches 40; Indels 98; Gaps 12;  
 Matches 66; Conservative 40; Mismatches 120; Indels 98; Gaps 12;

QY 149 IITPFHNNGCWSSSLRRTVHSVNRSPPEVIAEVVW-----DDFSREHLKPL 196
 Db 22 WVCTTSMRBYGPPADAVSVLAQTHDDL--EVILVUDGNDVVERVRAFDGFBHVVH 79
 QY 197 EDYMLALFESVSRVIRLTKERGLTRTRMIGASVAVGAVGDVITFLDSHCEANVNWLPFLDRAR 256
 Db 80 ND-----ENQGISYSRTKCKAKIASGDIVAFDDDATAEDWISQOLGVVAE 125
 QY 257 NRKTIIVCFCMDVIDHDFFRYETVQAGDAMRGAFDNEYMYKRPPIPPEQKADSDPFPV 316
 Db 126 TDAJAVGGS--DWDPD---WQTKPDPFPFETW-----LVGCVTPGFAADGEYTRN 171
 QY 317 MAGGLFAVDRKWWELGGYDPGLEIWGEGQEYTSFKWMCGGRMEDIPCSRVGHIYRCKV 376
 Db 172 TYGSNISFRDQFLDVGCGYDPNT-----GRKGD-----KHLQ 203
 QY 377 PYKVPAGTSLARNLKVRAEVWMDBEYAVIYQRREYVLSAGDVAVOKRLSSLNCKSPK 436
 Db 204 AHRAEPVGIRL-----RBEYGRCKNFT-----DARVHHLFDYRG--EFG 241
 QY 437 WFTPKIAND-----LPFPYPP 454
 Db 242 WLVSRSFHQGYSKRVMDLWPDKE 265

RESULT 24

S74669 hypothetical protein S111664 - *Synechocystis* sp. (strain PCC 6803)  
 C;Species: *Synechocystis* sp.  
 A;Variety: PCC 6803  
 C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
 C;Accession: S74669  
 R;Kaneko, T.; Sasano, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
 DNA Res. 3, 109-136, 1996  
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
 B;Reference number: S74669  
 A;Accession: S74669  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-312 <KAN>  
 A;Cross-references: UNIPROT:PT2806; EMBL:D90900; GB:AB001339; NID:91651768; PIDN:AA16821  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 4.1%; Score 135.5; DB 2; Length 312;  
 Best Local Similarity 22.7%; Pred. No. 0.0038; Mismatches 69; Indels 55; Gaps 14;  
 Matches 69; Conservative 51; Mismatches 129; Indels 55; Gaps 14;

```

QY 148 SIIIPPHNEGSSSLRRTVHSVNRSPPEVIA-----ETVWUDFSDREHLKKPLHDYALFP 204
Db 4 SSVIPTNR-LPLBKCGIRALEQQTDFETLISGYETIVVUDGSDTGTAW-LDGHCGFP 61
QY 205 SIIIPPHNEGSSSLRRTVHSVNRSPPEVIA-----ETVWUDFSDREHLKKPLHDYALFP 264
Db 62 HVTWHE-QDHOQPAARNLGVKAEGSDTITFSDLWTFDLOAHAGLORGOK---- 115
QY 265 MIDVIDHDFFRYETVQAGDAMRGAFDNEYMYKRPPIPPEQKADSDPFPFETW----- 324
Db 116 -----QNSDRVFTYGVNTQED-----SPQSERPKLTD-YSAFATGWA 159
QY 325 DRKWFWEIIGGYYGPGLEIWGEGQEYTSFKWMCGGRMEDIPCSR-VGHIYRKVYPKAG 383
Db 160 AKWNLLEGLFDGFOQWGEWDLBGLVRLKNIKLGL--VKCPRAVG--YHWHPP---- 210
QY 384 VSILARNLKVRAEVWMDBEYAVIYQRREYVLSAGDVAVOKRLSSLNCKSPK 442
Db 211 -SLOQTPKLDIBIORGMGVLFYQKHTW-----EVRLMIQMTRLHWFL-- 254
QY 443 AWDL 446
Db 255 -KGL 297

```

D87505 Glycooyl transferase family protein CC2077 [imported] - Caulobacter crescentus  
 C;Species: Caulobacter crescentus  
 C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
 C;Accession: D87505  
 R;Nieman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Hidelberg, J.  
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
 n, J.; Ermoljeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A;Title: Complete Genome Sequence of Caulobacter crescentus.  
 A;Reference number: A87249; MNUID:21173698; PMID:11259647  
 A;Accession: D87506  
 A;StatutB: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-318 <STO>  
 A;Cross-references: UNIPROT:09A6L6; GB:AE005673; NID:913423558; PIDN:AAK24048.1; GSPDB:G  
 C;Genetics:  
 A;Gene: CC2077

Query	Match	Score	DB	Length
QY	Query Match	4.1%		
Db	Best Local Similarity	24.7%	Pred. No.	0.0039
Matches	63;	Conservative	Mismatches	102;
Db	Indels	51;	Gap	10
QY	143 TLPENTSIIIPPFHNIGWSSLIRTRVHVNTRASPELVAEIVLVDFFSIREHKKPREDIVAL	2022		
Db	17 TRPRLSLVAVVWVMTG-PALMESIRHALDE--PRVDEFIIVNGSSLLADAAM-LRDARR	71		
QY	203 FPSTRILRTRKREGIRTRMLGASATGVDVITFLOSHCEANVNMLPP---LLDRTARNR	258		
Db	72 EPRTRLQSGLGNIGPARAANNGATAKGDDELVFLNDAA---PFTGATIAALRERAAID	126		
QY	259 KTIUCPMIDVIDHDDPFRYETOAGDMAQGAFD-----WEMYYKRIPI	298		
Db	127 PSCPVGARVNFNTGQ---TEBORGRRGEITPVTUTLISLSKLSATUPPLRFEIIRGEPE	182		
QY	299 IPEBLQKADPSPDPPFSPVMAGGLFFAVDRKWNWELGGYDPG---LEIWMGEQYVIFISKWMC	356		
Db	183 IPP-----FPVDTPTISGACFTVSARDQFQLSGPDEGYFLH---EDIDLWARRQ	231		
QY	357 GGRMDIPICSRVGHT	371		
Db	232 GGCFLRFQPHARVHIL	246		

Search completed: November 22, 2004, 13:44:12  
Job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: November 22, 2004, 13:38:22 ; Search time 67 seconds

(without alignments) 5178.372 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 328

Sequence: 1 MRRKEKRLLQGAVALVLAALV.....TQQMLFERTNSTVLEKPNRN 603

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : UniProt\_02,\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	3278	100.0	603	1	GLI10_HUMAN	Q86811	h polypepti
2	3157	96.3	603	1	GLI10_RAT	Q93577	r polypepti
3	3138	95.7	634	2	BAD27405	Bad21405	mus muscu
4	3137	95.7	603	1	GLI10_MOUSE	Q69879	m polypepti
5	3137	95.7	603	2	AAH60617	Aah60617	mus muscu
6	1404	42.8	644	2	Q70929	Q70929	anophales g
7	1392.5	42.5	622	1	GLI10_CAEEL	Q45947	caenorhabdi
8	1360	41.5	666	1	GLT2_DRONE	Q6wv16	drosophila
9	1360	41.5	666	2	Q8P6703	Aaq56703	drosophila
10	1341.5	40.9	599	2	Q7PZM5	Q7Pzwm5	anophales g
11	1247	38.0	1003	2	Q7D9R0	Q7D9r0	anophales g
12	1186	36.2	659	1	GLT4_DRONE	Q8i442	drosophila
13	1186	36.2	659	2	AAQ56701	Aaq56701	drosophila
14	1172.5	35.8	650	1	GLT9_DRONE	Q8Pc99	drosophila
15	1159	35.4	645	2	Q7QBT5	Q7Qbt5	anophales g
16	1144	34.9	518	2	Q7Q048	Q7Q048	anophales g
17	1125	34.3	559	1	GLT1_HUMAN	Q7Pzwm5	anophales g
18	1111	34.1	559	1	GLT1_PIG	Q8i442	drosophila
19	1116	34.0	559	1	GLT1_RAT	Q8P6703	drosophila
20	1115	34.0	559	1	GLT1_BOVIN	Q8Pc99	drosophila
21	1115	34.0	559	1	GLT1_MOUSE	Q7Q815	anophales g
22	1113	34.0	601	1	GLT7_CAEEL	Q7Q6048	anophales g
23	1112.5	33.9	556	2	Q6GM51	Q10472	h polypepti
24	1111	33.9	617	2	AAQ56702	Q29121	a polypepti
25	1109	33.8	556	1	GLI13_MOUSE	Q10473	r polypepti
26	1109	33.8	556	1	GLI13_RAT	Q07537	b polypepti
27	1109	33.8	556	2	AAQ75749	Q08912	m polypepti
28	1108.5	33.8	630	1	GLT5_DRONE	Q6wv17	drosophila
29	1106	33.7	556	1	GLI13_HUMAN	Q8iuc8	h polypepti
30	1087	33.2	559	2	Q6pa81	Q6pa81	xenopus lae
31	1087	33.2	559	2	AAH60419	Aah60419	xenopus lae

Result No.	Score	Query	Match	Length	DB	ID	Description
1	3278	100.0	603	1	GLI10_HUMAN	Q86811	h polypepti
2	3157	96.3	603	1	GLI10_RAT	Q93577	r polypepti
3	3138	95.7	634	2	BAD27405	Bad21405	mus muscu
4	3137	95.7	603	1	GLI10_MOUSE	Q69879	m polypepti
5	3137	95.7	603	2	AAH60617	Aah60617	mus muscu
6	1404	42.8	644	2	Q70929	Q70929	anophales g
7	1392.5	42.5	622	1	GLI10_CAEEL	Q45947	caenorhabdi
8	1360	41.5	666	1	GLT2_DRONE	Q6wv16	drosophila
9	1360	41.5	666	2	Q8P6703	Aaq56703	drosophila
10	1341.5	40.9	599	2	Q7PZM5	Q7Pzwm5	anophales g
11	1247	38.0	1003	2	Q7D9R0	Q7D9r0	anophales g
12	1186	36.2	659	1	GLT4_DRONE	Q8i442	drosophila
13	1186	36.2	659	2	AAQ56701	Aaq56701	drosophila
14	1172.5	35.8	650	1	GLT9_DRONE	Q8Pc99	drosophila
15	1159	35.4	645	2	Q7QBT5	Q7Qbt5	anophales g
16	1144	34.9	518	2	Q7Q048	Q7Q048	anophales g
17	1125	34.3	559	1	GLT1_HUMAN	Q7Pzwm5	anophales g
18	1111	34.1	559	1	GLT1_PIG	Q8i442	drosophila
19	1116	34.0	559	1	GLT1_RAT	Q8P6703	drosophila
20	1115	34.0	559	1	GLT1_BOVIN	Q8Pc99	drosophila
21	1115	34.0	559	1	GLT1_MOUSE	Q7Q815	anophales g
22	1113	34.0	601	1	GLT7_CAEEL	Q7Q6048	anophales g
23	1112.5	33.9	556	2	Q6GM51	Q10472	h polypepti
24	1111	33.9	617	2	AAQ56702	Q29121	a polypepti
25	1109	33.8	556	1	GLI13_MOUSE	Q10473	r polypepti
26	1109	33.8	556	1	GLI13_RAT	Q07537	b polypepti
27	1109	33.8	556	2	AAQ75749	Q08912	m polypepti
28	1108.5	33.8	630	1	GLT5_DRONE	Q6wv17	drosophila
29	1106	33.7	556	1	GLI13_HUMAN	Q8iuc8	h polypepti
30	1087	33.2	559	2	Q6pa81	Q6pa81	xenopus lae
31	1087	33.2	559	2	AAH60419	Aah60419	xenopus lae

### ALIGNMENTS

Result No.	Score	Query	Match	Length	DB	ID	Description
1	3278	100.0	603	1	GLI10_HUMAN	Q957J1	caenorhabdi
2	3157	96.3	603	1	GLI10_RAT	Q6wv24	xenopus lae
3	3138	95.7	634	2	BAD27405	Aah70527	xenopus lae
4	3137	95.7	603	1	GLI10_MOUSE	Q6wv21	xenopus tro
5	3137	95.7	603	2	AAH60617	Aah67317	xenopus tro
6	1404	42.8	644	2	Q70929	Q886f2	h sapien
7	1392.5	42.5	622	1	GLI10_CAEEL	Q8wv48	drosophila
8	1360	41.5	666	1	GLT2_DRONE	Aas256704	drosophila
9	1360	41.5	666	2	Q8P6703	Q6wv41	xenopus lae
10	1341.5	40.9	599	2	Q7PZM5	Aah71009	xenopus lae
11	1247	38.0	1003	2	Q7D9R0	Q6wv425	mus muscu
12	1186	36.2	659	1	GLT4_DRONE	Q6wv425	h sapien
13	1186	36.2	659	2	AAQ56701	Q6wv425	h sapien
14	1172.5	35.8	650	1	GLT9_DRONE	Q6wv419	h sapien
15	1159	35.4	645	2	Q7QBT5	Q6wv419	h sapien
16	1144	34.9	518	2	Q7Q048	Q6wv419	h sapien
17	1125	34.3	559	1	GLT1_HUMAN	Q6wv419	h sapien
18	1111	34.1	559	1	GLT1_PIG	Q6wv419	h sapien
19	1116	34.0	559	1	GLT1_RAT	Q6wv419	h sapien
20	1115	34.0	559	1	GLT1_BOVIN	Q6wv419	h sapien
21	1115	34.0	559	1	GLT1_MOUSE	Q6wv419	h sapien
22	1113	34.0	601	1	GLT7_CAEEL	Q6wv419	h sapien
23	1112.5	33.9	556	2	Q6GM51	Q10472	h polypepti
24	1111	33.9	617	2	AAQ56702	Q29121	a polypepti
25	1109	33.8	556	1	GLI13_MOUSE	Q10473	r polypepti
26	1109	33.8	556	1	GLI13_RAT	Q07537	b polypepti
27	1109	33.8	556	2	AAQ75749	Q08912	m polypepti
28	1108.5	33.8	630	1	GLT5_DRONE	Q6wv17	drosophila
29	1106	33.7	556	1	GLI13_HUMAN	Q8iuc8	h polypepti
30	1087	33.2	559	2	Q6pa81	Q6pa81	xenopus lae
31	1087	33.2	559	2	AAH60419	Aah60419	xenopus lae



QY	181	VJYDPSDREHKKPFLDYMALPSPVILRKREGILIRTMIGASSATGVDITELDS
Db	181	VVYDDPSDREHKKPFLDYMALPSPVILRKREGILIRTMIGASSATGVDITELDS
QY	241	EAVNNWLPPLDRIARURKTVCPMDIVIDHDIFRVEIQADAMRQAFDWEMYKIEP
Db	241	EAVNNWLPPLDRIARURKTVCPMDIVIDHDIFRVEIQADAMRQAFDWEMYKIEP
Db	301	PELQKADPSDPESPWAGGLEAVDGFNFWELGGYQGLETWGGBGEYELSKWMMGG
QY	301	PELQKADPSDPESPWAGGLEAVDGFNFWELGGYQGLETWGGBGEYELSKWMMGG
Db	361	EDJPCSRGHTKRVKPVTKPAGVSLARNLKVAEWMDAEAYVWQRPRYRHSAG
QY	361	EDJPCSRGHTKRVKPVTKPAGVSLARNLKVAEWMDAEAYVWQRPRYRHSAG
Db	421	AVVKKLRSLSINCKSPKQFMKTLAWDLKPFYPPVEPPRAWGSEBIRNGTGLCADTKGAA
QY	421	AVVKKLRSLSINCKSPKQFMKTLAWDLKPFYPPVEPPRAWGSEBIRNGTGLCADTKGAA
Db	481	SPLRLEGVRGRGEAAMNNMQVFTWREDIRPGDPQHTKCFDAlSHTSPVTLDDC
QY	541	MKGNLQWYKDKTILHVSSGQMDSESDARLIMPTQPSLTLQNLHETNSTLB
Db	541	MKGNLQWYKDKTILHVSSGQMDSESDARLIMPTQPSLTLQNLHETNSTLB
QY	601	NRN 603
Db	601	NRN 603

CC intermediate level in heart, brain, spleen, lung, stomach, cervix  
CC and uterus.

CC -!- DOMAIN: There are two conserved domains in the glycosyltransferase  
CC region: the N-terminal domain (domain A, also called GT1 motif),  
CC which is probably involved in manganese coordination and substrate  
CC binding and the C-terminal domain (domain B, also called  
CC Gal/GalNAc-T motif), which is probably involved in catalytic  
CC reaction and UDP-Gal binding (By similarity).

CC -!- DOMAIN: The ricin B-type lectin domain binds to GalNAc and  
CC contributes to the glycopeptide specificity (By similarity).

CC -!- SIMILARITY: Belongs to the glycosyltransferase family 2, GalNAc-T  
CC subfamily.

CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.

CC -!- CAUTION: According Ref.1, this enzyme is unable to transfer GalNAc  
CC onto serine or threonine residue on the protein receptor, but  
CC instead requires the prior addition of a GalNAc on a peptide  
CC before adding the required GalNAc moieties, thereby acting as a  
CC glycopeptide transferase.

CC -!- CAUTION: Was originally (Ref.1) termed GalnT9/PP-Galnase 9.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi-bib.ch/announce/>  
CC or send an email to license@ebi-bib.ch).

CC

DR EMBL; AF241241; ARKS5498.1; -.

DR HSUP; P26514; 1KNL.

DR RGD; 69409; Galnt10.

DR InterPro; IPR001173; Glyco\_trans\_2.

DR InterPro; IPR00897; Ricin\_B\_lectin.

DR Pfam; PF0035; Glyco\_transf\_2; 1.

DR Pfam; PF00652; Ricin\_B\_lectin; 3.

DR SMART; SM00458; RICIN\_BLECTIN; 1.

DR PROSITE; PS5231; RICIN\_BLECTIN; 1.

KW Calcium; Glycosyltransferase; Golgi stack; Lectin; Manganese;

KW Signal-anchor; Transmembrane.

FT DOMAIN 1 11 Cyttoplasmic (potential) -

FT DOMAIN 12 31 Signal-anchor for type II membrane

FT TRANSMEM 12 31 Protein (potential).

FT DOMAIN 32 603 Luminal (potential).

FT DOMAIN 144 253 Catalytic subdomain A..

FT DOMAIN 311 373 Catalytic subdomain B..

FT DOMAIN 458 590 Ricin B-type lectin.

FT DISULFID 471 488 By similarity.

FT DISULFID 523 538 By similarity.

FT DISULFID 563 578 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 124 124 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 146 146 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 593 593 N-linked (GlcNAc. . .) (Potential).

FT SEQUENCE 603 AA; 69116 MW; 194DEDA626A4BBF CRC64;

Query Match 96.3%; Score 3157; DB 1; Length 603;  
Best Local Similarity 96.0%; Pred. No. 2e-247;  
Matches 579; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

QY 1 MRRKEKRLIQQAVAVLVALVLLPVLPGVGLWLYREROPDGTFGGSGAAVAPAGQGSHSRK 60

Db 1 MRRKEKRLIQQAVAVLVALVLLPVLPGVGLWLYREROPDGTFGGSGAAVAPAGQGSHSRK 60

QY 61 KTFPLGGDKQKLDWDKEAIRRDAQRVGNGEQRGPRYPMTAERDAERDQAYRENGNIYVSDK 120

Db 61 KTFPLGGAEQKLDWDKEAIRRDAQRVGNGEQRGPRYPMTAERDAERDQAYRENGNIYVSDK 120

QY 121 ISLNSLSPDIRHPCNSKRYLETLNTSIIIPFNEGMSLRLTHSVNRSPELVAI 180

Db 121 ISLNSLSPDIRHPCNSKRYLETLNTSIIIPFNEGMSLRLTHSVNRSPELVAI 180

QY 181 LVVDDFSREHKKPLEYMAALPSVRLRKKEGRJTRMIGASANTGDVITFLSHC 240

Db 181 LVVDDFSREHKKPLEYMAALPSVRLRKKEGRJTRMIGASANTGDVITFLSHC 240

QY 241 BANVNWLPFLDLTARNRTIVCPMDIDHDPFRYETQAGDAMGAFDWEMYKRIP 300

Db 241 BANVNWLPFLDLTARNRTIVCPMDIDHDPFRYETQAGDAMGAFDWEMYKRIP 300

QY 301 PELOKADPSDPSPYPMAGGLFADVRKFWELGGYDPELUEINGEQEYVIEFKYMCGRM 360

Db 301 PELOKADPSDPSPYPMAGGLFADVRKFWELGGYDPELUEINGEQEYVIEFKYMCGRM 360

QY 361 EDIPCSRWHYIYRKVPPKYPAGVSLANLKRVAEVMWDEYAEYIYQRPPEYRHLASGDV 420

Db 361 EDIPCSRWHYIYRKVPPKYPAGVSLANLKRVAEVMWDEYAEYIYQRPPEYRHLASGDV 420

QY 481 SPRLLEGYVGRGDRBAAWNMNOVFTWBDIREGDPORTKCCFDSLHTSPVLYDCHS 540

Db 481 SPRLLEGYVGRGDRBAAWNMNOVFTWBDIREGDPORTKCCFDSLHTSPVLYDCHS 540

QY 541 MKGQIQLWKDKDQKLYHPPVSGSCMDCSSSDHRFMNTNPSSLTQOMIPREHTNSTVLEKF 600

Db 541 MKGQIQLWKDKDQKLYHPPVSGSCMDCSSSDHRFMNTNPSSLTQOMIPREHTNSTVLEKF 600

QY 601 NRN 603

Db 601 NRN 603

RESULT 3

BAD21405 PRELIMINARY; PRT; 634 AA.

AC BAD21405; 01-JUN-2004 (TREMblrel. 27, Created)

DT 01-JUN-2004 (TREMblrel. 27, Last sequence update)

DE MFJW00205; protein (Fragment).

GN MFJW00205.

OS Mus musculus (Mouse)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciuromathi; Muridae; Murinae; Mus.

OX NCBI\_TAXID10909;

RN [1]

RP SEQUENCE FROM N.A.

RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiracka S., Roga Y., Kitamura H., Nakagawa T., Negase T., Ohara O., Roga H.;

RT "Prediction of the Coding Sequences of 110 Mouse Homologues of PLJ Gene: The Complete Nucleotide Sequences of 110 Mouse PLJ-homologous cDNA Identified by Screening of Terminal Sequences of cDNA Clones Randomly Sampled from Size-Fractionated Libraries."

RT Submitted (PBB-2004) to the EMBL/Genbank/pdbJ databases.

RL EMBL; AKJ31155; BAD21405.1; -.

FT NON\_TER 1 1.

FT SEQUENCE 634 AA; 72315 MW; C79B82D0052C81 CRC64;

Query Match 95.7%; Score 3138; DB 2; Length 634;  
Best Local Similarity 95.5%; Pred. No. 7.6e-245;  
Matches 576; Conservative 9; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRRKEKRLIQQAVAVLVALVLLPVLPGVGLWLYREROPDGTFGGSGAAVAPAGQGSHSRK 60

Db 32 MRRKEKRLIQQAVAVLVALVLLPVLPGVGLWLYREROPDGTFGGSGAAVAPAGQGSHSRK 91

QY 61 KTFPLGGDKQKLDWDKEAIRRDAQRVGNGEQRGPRYPMTAERDAERDQAYRENGNIYVSDK 120

Db 92 KTFPLGGAEQKLDWDKEAIRRDAQRVGNGEQRGPRYPMTAERDAERDQAYRENGNIYVSDK 151

QY 121 ISLNSLSPDIRHPCNSKRYLETLNTSIIIPFNEGMSLRLTHSVNRSPELVAI 180

Db 152 ISLNSLSPDIRHPCNSKRYLETLNTSIIIPFNEGMSLRLTHSVNRSPELVAI 211

QY 181 VLVVDFSDREHHLKPKLEDMALPSVRLRTKREGGLRTTRQAGASVATGVDVTFUDSHC 240  
 QY 212 VLVVDFSPREHHLKPKLEDMALPSVRLRTKREGGLRTTRQAGASVATGVDVTFUDSHC 271  
 QY 241 EANVNWLPPLLDLARNRKTIVCVMIDVDDHDPFRYETOAGDAMGAFDWEWYKRPPIP 300  
 QY 272 EANVNWLPPLLDLARNRKTIVCVMIDVDDHDPFRYETOAGDAMGAFDWEWYKRPPIP 331  
 QY 301 PELQKADPSDPFPSPVAGLFAVDRKDFWELGCGYDGLBINGEYQISPKYWMCGRM 360  
 QY 332 PELQKADPSDPFPSPVAGLFAVDRKDFWELGCGYDGLBINGEYQISPKYWMCGRM 391  
 QY 361 EDIPCSRQVHIIYRKVVKVUPVAGLFAVDRKDFWELGCGYDGLBINGEYQISPKYWMCGRM 420  
 Db 392 EDIPCSRQVHIIYRKVVKVUPVAGLFAVDRKDFWELGCGYDGLBINGEYQISPKYWMCGRM 451  
 QY 421 AVOKKLRSILNCSPKWKWNTKIAWDLPKVPPYEPAAWGBTRNGVGLCAGTAKHGAIG 480  
 QY 452 VAOQKLKVLSINCKSPKWKWNTKIAWDLPKVPPYEPAAWGBTRNGVGLCAGTAKHGAIG 511  
 QY 481 SPURLEGCVGRGEAAMNMQVFTWEDRISQDPQHTRKKEPDAILSHTSPYLDCS 540  
 Db 512 SPURLEGCVGRGEAAMNMQVFTWEDRISQDPQHTRKKEPDAILSHTSPYLDCS 571  
 QY 541 MKGNOLWKYRKOKTLYHPVSGSCMDSCSDDHRYVPMTCNSIQQWLFHTNSTVLENP 600  
 Db 572 MKGNOLWKYRKOKTLYHPVSGSCMDSCSDDHRYVPMTCNSIQQWLFHTNSTVLENP 631  
 QY 601 NRM 603  
 Db 632 NKN 634

RESULT 4

GU10\_MOUSE STANDARD; PRT; 603 AA.

AC Q6P9S7; Q6KAQ2; Q8BZU8; Q91YJ6;  
 DT 01-OCT-2004 (Rel. 45, Created)  
 DT 01-OCT-2004 (Rel. 45, Last sequence update)  
 DE Peptidyl-N-acetylglactosaminyltransferase 10 (EC 2.4.1.41)  
 DE (Protein-UDP acetylglactosaminyltransferase 10) (UPP-GalNAc-Polypeptide  
 DE GalNAc-Polypeptide N-acetylgalactosaminyltransferase 10) (Polypeptide  
 DE GalNAc transferase 10) (GalNAc-T10) (pp-GalTase 10).

GN Name=GalNT10;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TAXID=10909;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hirakawa S.,  
 RT Soga Y., Kitamura H., Nakadawa T., Nagase T., Ohara O., Koga H.;  
 RT "Prediction of the coding sequences of mouse homologues of *Flj* genes:  
 RT the complete nucleotide sequences of 110 mouse *FLJ*-homologous cDNAs  
 RT identified by screening of terminal sequences of cDNA clones randomly  
 RT sampled from size-fractionated libraries.";  
 RL DNA Res. 11:167-180(2004).  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6, and FVB/N; TISSUE=Brain, and Breast tumor;  
 RX MEDLINE=22308257; PubMed=1477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derg J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemesh C.M., Schuler G.D.,  
 RA Aitschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Blatt N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., May S.I., Wang J., Haleh P.,  
 RA Blatchko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
 RA Brownstein M.J., Usdin T.B., Tobiishi Y., Carninci P., Prange C.,  
 RA Raha S.S., Loqueland N.A., Peters G.J., Abramson R.D., Mulany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huljk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rooriger A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smialius D.E.,  
 RA Scherzer A., Schain J.E., Jones S.J.M., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences."; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RL [3]  
 RP SEQUENCE OF 76-603 FROM N.A.  
 RC SPRAIN-C57BL/6J, TISSUE=Colon;  
 RX MEDLINE=22304683; PubMed=1466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nakaiko I., Osato N., Saito R., Suzuki H., Yamamoto I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Rume D.A., Quackenbush J.W.,  
 RA Schriml L.M., Karapin A., Matsuda H., Batyalov S., Beisel K.W.,  
 RA Blake J.A., Bradd D., Brusick V., Chonkla C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gasterlend T., Garibaldi M., Gissi C., Godik A., Gough J.,  
 RA Grimmond S., Gusickich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawai H., Kawabwa Y., Kozdiarski R.M., King J.B.L.,  
 RA Konagai A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Magliott D.R., Maltaias L., Marchionni L., McKenlie B., Min I.H.,  
 RA Nagashima T., Numata K., Okido T., Pavon W.J., Perreia G., Pesole G.,  
 RA Petrovsky N., Hillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed D.J., Reed J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultan A., Takeaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynnshaw-Boris A., Yang S., Yang T., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmo A., Carnici P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Saito K.,  
 RA Shizaki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume M., Itohaki K., Itoh M., Kogawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibaoka K., Shingawa A.,  
 RA Yasunishi A., Yoshimoto M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashita Y.,  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT full-length cDNAs."; Nature 420:563-573(2002).  
 RN [4]  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=21264503; PubMed=11278534; DOI=10.1074/jbc.M009638200;  
 RA Ten Hagen K.G., Bedi G.S., Tetaert D., Kingsley P.D., Hagen F.K.,  
 RA Balsys M.M., Beres T.M., Deacon P., Tabak L.A.;  
 RT "Cloning and characterization of a ninth member of the UPP-  
 RT GalNAc-Polypeptide N-acetylglactosaminyltransferase family,  
 RT ProGantase-Ts."; J. Biol. Chem. 276:17395-17404(2001).  
 CC -I- FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has activity toward Muc5AC and EA2 peptide  
 CC substrates (By similarity).  
 CC -I- CATALYTIC ACTIVITY: UPP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -I- COFACTOR: Manganese and calcium (By similarity).  
 CC -I- PATHWAY: Glycosylation.  
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
 CC similarity).  
 CC -I- TISSUE SPECIFICITY: Expressed at higher level than GalNT9. In the  
 CC developing hindbrain region of E14.5 embryos it accumulates in the  
 CC rapidly dividing, undifferentiated ventricular zone adjacent to  
 CC the pons. It also accumulates in the regions immediately rostral  
 CC and caudal to the dorsal rhombic lips differentiating into the  
 CC cerebellum. Not expressed in the developing choroid plexus.  
 CC -I- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called GT1 motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC Gal/GalNAc-T motif), which is probably involved in catalytic

reaction and UDP-Gal binding (By similarity).  
 CC reaction and UDP-Gal binding (By similarity).  
 CC -|- DOMAIN: The ricin B-type lectin domain binds to GalNAc and  
 CC contributes to the glycopptide specificity (By similarity).  
 CC -|- SIMILARITY: Belongs to the glycosyltransferase family 2, GalNAc-T  
 subfamily.  
 CC -|- CAUTION: Contains 1 ricin B-type lectin domain.  
 CC -|- CAUTION: According to experiments made in rat, this enzyme is  
 unable to transfer GalNAc onto serine or threonine residue on the  
 protein receptor but instead requires the prior addition of a  
 GalNAc on a peptide before adding additional GalNAc moieties,  
 thereby acting as a glycopeptide transferase.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC DR MGD; MGJ:11890480; Galnt10.  
 CC DR GO:0004633; P-polyPeptide N-acetylgalactosaminyltransferase . . . IDA.  
 CC DR GO:0006493; P-O-linked glycosylation; IDA.  
 CC DR SMART; SM00458; RICIN\_B\_LECTIN; 1.  
 CC DR PROSITE; SS0231; RICIN\_B\_LECTIN; 1.  
 CC DR Calcium, Glycosyltransferase; Golgi stack; Lectin; Manganese;  
 KW Signal-anchor; Transferase; transmembrane.  
 CC DR DOMAIN 1 1 Cyttoplasmic (potential).  
 CC DR TRANSMEM 12 31 Signal-anchor for type II membrane  
 protein (potential).  
 CC FT DOMAIN 32 603 Luminal (potential).  
 CC FT DISULFID 458 590 Catalytic subdomain A.  
 CC FT DISULFID 471 488 Catalytic subdomain B.  
 CC FT DOMAIN 144 253 Ricin B-type lectin.  
 CC FT DISULFID 523 538 By similarity.  
 CC FT DISULFID 563 578 N-linked (GalNAc. . .) (potential).  
 CC FT CARBOHYD 124 124 N-linked (GalNAc. . .) (potential).  
 CC FT CARBOHYD 146 146 N-linked (GalNAc. . .) (potential).  
 CC FT CARBOHYD 593 593 N-linked (GalNAc. . .) (potential).  
 CC FT CONFLICT 233 233 V -> T (in Ref. 1).  
 CC SQ SEQUENCE 603 AA; 69116 MW; FPP5FBATE1DD7544 CRC64;

Query Match 95.7%; Score 3137; DB 1; Length 603;  
 Best local Similarity 95.4%; Pred. No. 8-6e-246;  
 Matches 515; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRRKKEKLQIAVALVALAATLPLPGLWALXEROPDGPSPGSGSRAVAPAGCOSHRSOK 60  
 QY 1 MRRKKEKLQIAVALVALAATLPLPGLWALXEROPDGPSPGSGSRAVAPAGCOSHRSOK 60  
 QY 1 MRRKKEKLQIAVALVALAATLPLPGLWALXEROPDGPSPGSGSRAVAPAGCOSHRSOK 60  
 QY 61 KTFPLGDKQKLKDMDKEATRDRDQRGVNGQGRGPMDAERDQAYRNGPNIVSDK 120  
 QY 61 KTFPLGDKQKLKDMDKEATRDRDQRGVNGQGRGPMDAERDQAYRNGPNIVSDK 120  
 QY 121 ISLNISLPLDIRPHNCNSKRVLETLNTSIIIPFENEGWGSLLRTTHSVIARSPPELVAI 180  
 QY 121 ISLNISLPLDIRPHNCNSKRVLETLNTSIIIPFENEGWGSLLRTTHSVIARSPPELVAI 180  
 QY 121 ISLNISLPLDIRPHNCNSKRVLETLNTSIIIPFENEGWGSLLRTTHSVIARSPPELVAI 180  
 QY 181 VLVDPSDREHLLKQDLEDYMLPPSPVRLTAKKRGGLRTRMLGASVATCDVITPLDSH 240  
 QY 181 VLVDPSDREHLLKQDLEDYMLPPSPVRLTAKKRGGLRTRMLGASVATCDVITPLDSH 240  
 DB EMBL; BC060617; AAC60617.1; -.

QY 241 EANVNLPLPLDITARNKTIQPMIDIDHDPFREYQAGDMRGADWEMYKR\_PIP 300  
 DB 241 EANVNLPLPLDITARNKTIQPMIDIDHDPFREYQAGDMRGADWEMYKR\_PIP 300  
 QY 301 PELOKADSDPFSPPVPMAGGLFAVDRKFWLGGYDPELEIINGEQEYBISFKWMMGGRM 360  
 DB 301 PELOKADSDPFSPPVPMAGGLFAVDRKFWLGGYDPELEIINGEQEYBISFKWMMGGRM 360  
 QY 361 EDIPCSRQHIVRKVVPKVPAGVSLANLKRVAEVMDYBASYVYORPPEYRHSACDV 420  
 DB 361 EDIPCSRQHIVRKVVPKVPAGVSLANLKRVAEVMDYBASYVYORPPEYRHSACDV 420  
 DB 481 SPLRLEGCVGRGERAWNNMQVFTWBDIDRGPDTKCFDALSHTSPTLYDCHS 540  
 DB 481 SPLRLEGCVGRGERAWNNMQVFTWBDIDRGPDTKCFDALSHTSPTLYDCHS 540  
 QY 541 MKGNOLWIKYRKDKTLYHPPVSGSCMDSCSDHRTFMNTCPSSITQOMWTFENTSTVLRK 600  
 DB 541 MKGNOLWIKYRKDKTLYHPPVSGSCMDSCSDHRTFMNTCPSSITQOMWTFENTSTVLRK 600

QY 601 NNR 603  
 DB 601 NNR 603

RESULT 5  
 ID AAH60617 PRELIMINARY PRT; 603 AA.  
 ID AAH60617, AC  
 DT 12-MAY-2004 (TREMblrel. 27, Created)  
 DT 12-MAY-2004 (TREMblrel. 27, Last sequence update)  
 DT 12-MAY-2004 (TREMblrel. 27, Last annotation update)  
 DE UDP-N-acetyl-alpha-D-galactosamine:polypeptide  
 DE N-acetylgalactosaminyltransferase 10.  
 RN NCBITaxonID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6; TISSUE=Brain;  
 RC MEDLINE=2238257; Published=1247932;  
 RA Struberg R. L., Feingold E. A., Grouse L. H., Derge J. G., Schuler G. D.,  
 RA Klausner R. D., Collins F. S., Wagner L. L., Shenmen C. M., Schaefer C. P., Blat N. K.,  
 RA Altschul S. F., Zeeberg B., Buetow K. H., Schaefer C. P., Blat N. K.,  
 RA Hopkins R. F., Jorcan H., Moore T., Max S. I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marsina K., Farmer A. A., Rubin G. M., Hong L.,  
 RA Stopleton M., Soares M. B., Bonaldo M. F., Rabkin T. L., Scheetz T. P.,  
 RA Stopleton M., Soares M. B., Bonaldo M. F., Rabkin T. L., Scheetz T. P.,  
 RA Raha S. S., Loquellano N. A., Peters G. J., Abramson R. D., Mullany S. J.,  
 RA Bosak S. A., McEwan P. J., McKernan J. H., Malek J. A., Gunaratne P. H.,  
 RA Richards S., Worley K. C., Hale S., Garcia A. M., Gay L. J., Hulyk S. W.,  
 RA Villalon D. K., Munoz D. M., Sodergren E. J., Lu X., Gibbs R. A.,  
 RA Fahey J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Blakesley R. W., Touchman J. W., Green B. D., Dickson M. C.,  
 RA Rodriguez J. J., Schmutz J., Myers R. M., Butterfield Y. S.,  
 RA Krzywinski M. I., Skalska U., Smalius D. E., Scherzer A., Schein J. E.,  
 RA Jones S. J., Marra M. A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6; TISSUE=Brain;  
 RA Strauberg R.;  
 RL Submitted (Oct-2003) to the EMBL/GenBank/DBJ databases.  
 EMBL; BC060617; AAC60617.1; -.

KW Transfase.  
 SQ SSEQUENCE 603 AA; 69116 MW; PPF55FBATE1DD7544 CRC64;  
 Query Match 95.7%; Score 3137; DB 2; Length 603;  
 Best Local Similarity 95.4%; Pred. No. 8.6e-246;  
 Matches 575; Conservative 10; Mismatches 18; Indels 0; Gaps 0;  
 QY 1 MRRKEKRLQIAVALVLAALVLLPVGWALYRERQPDGTPGSSGAAVPAACGSHSRQK 60  
 1 MRRKEKRLQIAVALAALVLLPVGWALYRERQPDGSPGGLGAAGAAVAAVOBLSRQK 60  
 Db 61 KTFELGSGOKLQKOMWNEKAIIRDQARQVNGEQRSPYPMDAEVDQARNGNIVSDK 120  
 61 KTFELGSGOKLQKOMWNEKAIIRDQARQVNGEQRSPYPMDAEVDQARNGNIVSDK 120  
 QY 121 ISJNRSLPDIRHPCNSKRYETLPLNTSIIIPFHNEGNSSLRVTHSVANRSPPELBARI 180  
 121 ISJNRSLPDIRHPCNSKRYETLPLNTSIIIPFHNEGNSSLRVTHSVANRSPPELBARI 180  
 Db 181 VLVDDFSREHLKPKLEDMPLFSVRILTRKREGILTRMIGASAATGDDVTFPLSHC 240  
 181 VLVDDFSREHLKPKLEDMPLFSVRILTRKREGILTRMIGASAATGDDVTFPLSHC 240  
 QY 241 EANVNLPLLDIARNRTIVCIMIDVHDPERYETQAGDAMGADWEMTYKRIP 300  
 241 EANVNLPLLDIARNRTIVCIMIDVHDPERYETQAGDAMGADWEMTYKRIP 300  
 Db 301 PELOKADSDFPFSPPVMAGGLFAVDRKWFELLGYPGEIWMGCBQEYTSFKWVMCGRM 360  
 301 PELOKADSDFPFSPPVMAGGLFAVDRKWFELLGYPGEIWMGCBQEYTSFKWVMCGRM 360  
 Qy 361 EDIPCSRYGHVYRKYKPVKPGASLNRKVAEWNDEAYEYIYORPEYRHLHSADY 420  
 361 EDIPCSRYGHVYRKYKPVKPGASLNRKVAEWNDEAYEYIYORPEYRHLHSADY 420  
 Db 421 AVOKKLRSSLNCISFKWMTKIAWLDLKYPUPPERAAMGEERNVGLCAKTHGALG 480  
 421 AVOKKLRSSLNCISFKWMTKIAWLDLKYPUPPERAAMGEERNVGLCAKTHGALG 480  
 Db 481 SPURLEGCVRGREAAWNMMQVFTFWREDIREDPQHKKFCDAIHTSPVTLYDCHS 540  
 481 SPURLEGCVRGREAAWNMMQVFTFWREDIREDPQHKKFCDAIHTSPVTLYDCHS 540  
 Db 481 SPURLEGCVRGREAAWNMMQVFTFWREDIREDPQHKKFCDAIHTSPVTLYDCHS 540  
 481 SPURLEGCVRGREAAWNMMQVFTFWREDIREDPQHKKFCDAIHTSPVTLYDCHS 540  
 Qy 541 MKGNOLWQYRKDKTLYHPVSGSCMDSCSBDHRIFMNTCPSLJTOQWLFHNTVLF 600  
 541 MKGNOLWQYRKDKTLYHPVSGSCMDSCSBDHRIFMNTCPSLJTOQWLFHNTVLF 600  
 Db 601 NRM 603  
 601 NRM 603  
 Db 601 NKN 603  
 601 NKN 603

RESULT 6

Q7Q0E9 PRELIMINARY; PRT; 644 AA.  
 ID Q7Q0E9  
 AC Q7Q0E9;  
 DT 01-MAR-2004 (TREMBREL 26, Created)  
 DT 01-MAR-2004 (TREMBREL 26, Last sequence update)  
 DE AgCP9480 (Fragment)  
 GN Name=agCG54007; ORFNames=ENSANGG00000009226;  
 OS Anopheles gambiae ssp. PEST.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.  
 OC NCBI\_TaxID:180454;  
 RN [1] Sequence from N.A.  
 RP STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.

RESULT 7

GLJ\_0\_CAEEL STANDARD; PRT; 622 AA.  
 ID GLJ\_0\_CAEEL  
 AC 05947;  
 DT 01-OCT-2004 (Rel. 45, Created)  
 DT 01-OCT-2004 (Rel. 45, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Putative polypeptide N-acetylgalactosaminyltransferase 10  
 DE (EC 2.4.1.41) (Protein-UDP acetylgalactosaminyltransferase 10) (UDP-  
 DE GalNAc-polypeptide N-acetylgalactosaminyltransferase 10) (pp-Gantase  
 DE 10).  
 GN Name=gly-10; ORFNames=Y45F10D.3;  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderrinae; Caenorhabditis.

OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B-1301 N2;  
 RA McMurray A.A.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: May catalyze the initial reaction in O-linked  
 oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 galactosamine residue to a serine or threonine residue on the  
 protein receptor (By similarity).  
 CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -!- COFACTOR: Manganese and calcium (By similarity).  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
 similarity).  
 CC -!- DOMAIN: There are two conserved domains in the glycosyltransferase  
 region: the N-terminal domain (domain A, also called Grn motif),  
 which is probably involved in manganese coordination and substrate  
 binding and the C-terminal domain (domain B, also called  
 Gal/GalNAc-T motif), which is probably involved in catalytic  
 reaction and UDP-Gal binding (By similarity).  
 CC -!- DOMAIN: The ricin B-type lectin domain binds to GalNAc and  
 contributes to the glycoprotein specificity (By similarity).  
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
 subfamily.  
 CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the European Bioinformatics Institute and the EMBL outstation -  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; AL021992; CAA16378.1; ALT\_INIT.  
 DR PIR; T26930; T26930.  
 DR WormPep; Y45F10D.3; CE16642.  
 DR InterPro; IPR00173; Glyco\_trans\_2.  
 DR InterPro; IPR00897; Ricin\_B\_Like.  
 DR InterPro; IPR00772; Ricin\_B\_lectin.  
 DR Pfam; PF00535; Glycosyltransferase\_2; 1.  
 DR Pfam; PF00532; Ricin\_B\_lectin; 3.  
 DR SMART; SM00438; RICIN\_1.  
 DR PROSITE; PS50231; RICIN\_B\_LCINTIN\_1.  
 KW Calcium; Glycosyltransferase; Golgi stack; Hypothetical protein;  
 Lectin; Manganese; Signal-anchor; Transferase; Transmembrane.  
 FT DOMAIN 1 12 Cyttoplasmic (Potential).  
 FT TRANSMEM 13 32 Signal-anchor for type II membrane  
 Protein (Potential).  
 FT DOMAIN 33 622 Luminal (Potential).  
 FT DOMAIN 156 268 Catalytic subdomain A.  
 FT DOMAIN 324 386 Catalytic subdomain B.  
 FT DOMAIN 519 622 Ricin B-type Lectin.  
 FT DISULFID 532 549 By similarity.  
 FT DISULFID 575 591 By similarity.  
 FT CARBOHYD 136 136 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 170 170 N-linked (GlcNAc. . .) (Potential).  
 SQ 622 AA; 7181 MW; FB3A2159C1B452 CRC64;

Query Match 42.5%; Score 1392.5; DB 1; Length 622;  
 Best Local Similarity 48.8%; Fred. No. 4.5e-104;  
 Matches 279; Conservative 87; Mismatches 175; Indels 31; Gaps 14;

QY 49 PAAAGGSHSRQKKTFLGLDGQQLKWDKEAARRDAQRVNGEQRPPYMDAERVDQ-- 106  
 59 PPAALGDEALDPPEKRGH-EK-K-WEDAEAYEKERREGGEWKGPKVPLPDKVEKEA 116

QY 107 -AYHENGKIVYKSKISLRSRPLPRHWNKNSKEYLETNTSTTIPFENREGSSLRT 164  
 117 LSLYKANGYKAVYKISLRSRPLPRHWNKNSKEYLETNTSTTIPFENREGSSLRT 176

QY 165 VHSLVLRSPRELVEIIVVUDFSDREHKKPLDYM--ALFSVIRTRKREGLRT 221  
 177 VYSVNRSPPELKEILIDVNDSEBPKALRQPLDFKLKNKIVKVKRKGGLIGR 236  
 DB 222 MLGASVATGDIVPLDSCBANVNLPLPUDRJARNTIVCPCMDVDDHDPRTETQG 281  
 DB 237 QLGQDAGTBILLFLDAMSEANTNWLPLPUDPFAEDYVIVVCFVFDV-DCETVPRQ-D 295  
 QY 282 DAMRGAFFDNEMYKRIPIPPELQKADPDPFESVMAGLFAYDRKWWELGGDYPGIBI 341  
 DB 296 EGARQSFNAFNPVTKLPLT-K-DRESPKPFNSPVMAGSGYFAISAKWVWELGGYDEGIDI 354

QY 342 WGGQYEVSPFKVNGGGRMEDICSRVHRYR-KYVPK-VPGVSLRNLKVEAWND 399  
 DB 355 WGGQYEVSPFKVNGGGRMEDICSRVHRYR-KYVPK-VPGVSLRNLKVEAWND 414

QY 400 EYAEYIKRPRKPHLSSGDAVQKLUSSLNCKSKPKMTCIAMLDRKFPVVEPPAA 459  
 DB 415 DYKETPLYKERRPGVGNADGLKLUKGIGRQLQKSFDFEMKELAFDQPKYVAPVKSA 474

QY 460 WGETRNVGTLGACDTHGAGLGSPLRLEGVR---GREAAWNMQVITFTWBRDIREGD 515

DB 475 EGEIRNNGINFCIDTQPKEQNQFRGLRKCTSDDKGGE----QDIRLIRWHDIPR-- 526

QY 516 PQHTRKFCDAISI---SPVTLQDHSKMGKQNLQWYR-KDQTLXHPVGSQMCSESQH 572

DB 527 --KGKICFDCSTSVDKAPVLFDPDKHSWKGQNLQFQYRAQKQIYHPISQCLTADENGK 584

QY 573 -IFMNTCINSSLTQMLFHTNSTVLEKFRN 603  
 DB 585 FLHMKKCDSSDQKQWATVDNELLERQAN 616

RESULT\_B  
 GLT6\_DRONE  
 ID GLT6\_DRONE STANDARD PRT; 666 AA.  
 AC 06W16; 095R40; 05V2X5;  
 DT 01-OCT-2004 (Rel. 45, Created)  
 DT 01-OCT-2004 (Rel. 45, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE N-acetylgalactosaminyltransferase 6 (EC 2.4.1.-) (Protein-UDP  
 DE acetylgalactosaminyltransferase 6) (pp-GalNAc-Polypeptide N-  
 DE acetylgalactosaminyltransferase 6) (pp-GalNAcase 6).  
 OS Name=P-gant6; ORGName=cc2103;  
 OC Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Notoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriidae; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A., ENZYME ACTIVITY, TISSUE SPECIFICITY, AND  
 DEVELOPMENTAL STAGE.  
 RC STRAIN=Colony-S; TISSUE=Embryo;  
 RX MEDLINE=2284110; PubMed=12029714; DOI=10.1074/jbc.M303836200;  
 RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein B.S., Zhang Z.;  
 RT "Functional characterization and expression analysis of members of the  
 UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family from  
 Drosophila melanogaster."  
 RT J. Biol. Chem. 278:35039-35048 (2003).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Berkeley;  
 RC MEDLINE=019606; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celtnar S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amanatides P.G., Scheer S.E., Li P.W., Hosking R.A., Galle R.F.,  
 George R.A., Lewin S.E., Richards S., Ashburner M., Hellenbourn S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazquez R.G., Chapple M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfaffkoch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Beno P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA	Buttis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	CC
RA	Cherry J.M., Dawley S., Dahle C., Davenport L.B., Davies P.,	CC
RA	de Pablo B., Decher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	CC
RA	Dodson K., Douc L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	CC
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	CC
RA	Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,	CC
RA	Gloedek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	CC
RA	Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,	CC
RA	Hottstein D., Houston K.A., Howland J.M., Wei M.-H., Ibsengawm C.,	CC
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	CC
RA	Kammel B.E., Kodira C.D., Kraft C., Kravitz S., Kulip D., Lai Z.,	CC
RA	Laskin P., Lei Y., Levinsky A.A., Li J.-H., Li Z., Liang Y., Lin X.,	CC
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	CC
RA	Merkulov G., Milshina N.V., Mobarrey C., Morris J., Mosheroff A.,	CC
RA	Mount S.M., Moy B., Murphy L., Muzny D.M., Nelson D.L.,	CC
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,	CC
RA	Palaiazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	CC
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,	CC
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,	CC
RA	Spieler E., Spradling A.C., Stapleton M., Strong R., Sun B.,	CC
RA	Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,	CC
RA	Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,	CC
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,	CC
RA	Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,	CC
RA	Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.,	CC
RA	"the genome sequence of <i>Drosophila melanogaster</i> ,"	CC
RN	Science 287:2185-2195 (2000).	CC
RP	SEQUENCE FROM N.A.	CC
RP	STRAIN-Berkeley; TISSUE=Embryo;	CC
RP	Middleton-22426066; PubMed=12237569;	CC
RA	Stapleton M., Carlson J.W., Brokstein P., Yu C., Champa M.,	CC
RA	George R.A., Guarini H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,	CC
RA	Rubin G.M., Celinkin S.E.,	CC
RT	"A <i>Drosophila</i> full-length cDNA resource."	CC
RL	Genome Biol. 3:RESEARCH0080-1-RESEARCH0080-8 (2002).	CC
RP	SEQUENCE FROM N.A.	CC
RP	STRAIN-Berkeley; TISSUE=Embryo;	CC
RP	Middleton-22426066; PubMed=12237569;	CC
RA	Stapleton M., Carlson J.W., Brokstein P., Yu C., Champa M.,	CC
RA	George R.A., Guarini H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,	CC
RA	Rubin G.M., Celinkin S.E.,	CC
RT	"A <i>Drosophila</i> full-length cDNA resource."	CC
RL	Genome Biol. 3:RESEARCH0080-1-RESEARCH0080-8 (2002).	CC
-1- FUNCTION: Glycopeptide transferase involved in O-linked oligosaccharide biosynthesis, which catalyzes the transfer of an N-acetyl-D-galactosamine residue to an already glycosylated peptide. In contrast to other proteins of the family, it does not act as a peptide transferase that transfers GalNAc onto serine or threonine residue on the protein receptor, but instead requires the prior addition of a GalNAc on a peptide before adding additional GalNAc moieties. Some peptide transferase activity is however not excluded, considering that its appropriate peptide substrate may remain unidentified. Prefers the diglycosylated Mu5AC-3/13 as substrate.	CC	
-1- COTACTOR: Manganese and calcium (By similarity).	CC	
-1- PATHWAY: Glycosylation.	CC	
-1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By similarity).	CC	
-1- TISSUE SPECIFICITY: In embryos, it is specifically expressed in the salivary glands from stage 12, becoming stronger at stage 13. Not expressed in other tissue. Further expressed during oogenesis, in the somatically derived follicle cells that surround the developing oocyte, which are involved in the maturation of the oocyte and construction of the egg shell, as well as playing a role in subsequent embryonic pattern formation.	CC	
-1- DEVELOPMENTAL STAGE: Expressed throughout embryonic, larval, pupal and adult stages, with increasing levels during larval development.	CC	
-1- DOMAIN: There are two conserved domains in the glycosyltransferase region: the N-terminal domain (domain A, also called GM1 motif), which is probably involved in manganese coordination and substrate binding and the C-terminal domain (domain B, also called Gal/GalNAc-T motif), which is probably involved in catalytic reaction and UDP-Gal binding (By similarity).	CC	
-1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and contributes to the glycopeptide specificity (By similarity).	CC	
-1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T subfamily.	CC	
CC	or send an email to license@ibb-sib.ch).	CC
EMBL: AY768067; ARQ56703; 1.	CC	
EMBL: AE003476; ARX47690; 1.	CC	
EMBL: AY661629; AAL29177; 1.	CC	
InterPro: IPR00897; Ricin_B like.	CC	
InterPro: IPR00772; Ricin_B lectin.	CC	
PFAM: PF00535; Glycos_transf_2; 1.	CC	
SMART: SH0456; Ricin_B lectin; 3.	CC	
PROSITE: SS50331; RICIN_B LECIN; 1.	CC	
KW	CALCIUM; Glycosyltransferase; Golgi stack; Lectin; Manganese; Signal-anchor; Transferase; Transmembrane.	CC
KW	Signal-anchor; Transferase; Transmembrane.	CC
FT	TRANSMEM 12 31	CC
FT	DISULFID 531	CC
FT	DISULFID 577 594	CC
FT	DISULFID 621 636	CC
FT	CARBOHYD 181 311	CC
FT	CARBOHYD 285 429	CC
FT	CARBOHYD 651 651	CC
FT	CARBOHYD 657 657	CC
FT	CONFFLICT 95 96	CC
FT	CONFFLICT 107 107	CC
FT	Q -> R (in Ref. 1 and 3).	CC
FT	E -> K (in Ref. 3).	CC
FT	SEQUENCE 499 499	CC
FT	666 AA; 76972 MW; CABCAC6CB4860600C CRC64;	CC
Query	Match 41.5%; Score 1360, DB 1, Length 666;	CC
Best	Local Similarity 49.0%; Fred. No. 2.2e-101;	CC
Matches	277; Conservative 76; Mismatches 160; Indels 52; Gaps 15;	CC
QY	67 DQGKUKDWHIDKEAIRDADQRGNGSGQRGPYPTMDAERVDOAYR--ENGNIVYSDKISL 123	CC
Db	121 DASVKKDWHIDYTFMEKDAKGVGLGEGKAKSTDQSDRDLKMSLENGNALLSDSIV 180	CC
QY	124 NRSLDQDIRHNCSKRYLETPNTSIIIPHEGNESGLRTVHSTANRSPBPELBVILV 183	CC
Db	181 NRSVSDIIRHPKCKRYVAKLPVSVILKNEVYLVHSVLNSRPELBMKILV 240	CC
QY	184 DDPDSREHKKPLEYMA_LFSPVILRTKREGLIRTRMAGASVATGDTPLSHCEA 242	CC
Db	241 DDPDSREHKKPLEYMA_LFSPVILRTKREGLIRTRMAGASVATGDTPLSHCEA 300	CC
QY	243 NNVWPLDPLDRIARNRKTVCPMIDVIDHDPPRYETOAGDAMRGAFDWMENYKRPITPE 302	CC
Db	301 NYNWLPPLPPEFALNKRATAVCPFPIDVHDHNFHRAQ-DEGARGAFDWEFPPYKRLPLP 359	CC
QY	303 LQKADPSDPESPWVAGGLAVDREWFLWEGGYDGELEWGEQEISFWMMCGRMED 362	CC
Db	360 DLK_RHADPFPKSPIMAGGLFAISRFPEWFLWEGYDGELEWGEQEISFWMMCGRMED 418	CC
QY	363 ICPCSRQVHRY----KIVPKVPGVSLARNLKEVARWMDVDEKAYITKORPE_YRHL 416	CC
Db	419 APCSRQHRYGPRHNPSPRK--GDUYHKNYKRVAVEMDEKNYLLSHGDGLYESVD 475	CC
Db	417 AGDVAVKKRSLSNCKSFKMFMTKIAWDLPLKFYPPVPEPAAMGEIRAVGTYTGLCADT- 474	CC
Db	476 FGDLTBQKAIRTKLNUCKSFKMFMEVAFLDMLKTPPVDPSPYAMGALONVGNONCLDTL 535	CC

QY	536	GRKKHKGKGMACADNIKTPOR-----TOPFELSWKRDLR---LRRKE
Db	536	GRKKHKGKGMACADNIKTPOR-----TOPFELSWKRDLR---LRRKE
QY	523	CFDA--ISHTSPVTLYDCHSMKGQNLWCKY-RKQKTLHVPVSG-SCMDCSESDHRIFMTC
Db	523	CFDA--ISHTSPVTLYDCHSMKGQNLWCKY-RKQKTLHVPVSG-SCMDCSESDHRIFMTC
QY	578	CDVQIWDANAPYPTWMLWDCHSQGNOQWYDVRHKQLKGTEGRROLELLPPSQEAVNKC
Db	578	CDVQIWDANAPYPTWMLWDCHSQGNOQWYDVRHKQLKGTEGRROLELLPPSQEAVNKC
QY	579	NPSSLTQWLFPEHTNSTVLEKERN 603
Db	579	NPSSLTQWLFPEHTNSTVLEKERN 603
QY	637	DIDNRFOQWNFGSPNKTALDNYSQD 661
Db	637	DIDNRFOQWNFGSPNKTALDNYSQD 661
RESULT 9		
AAQ56703	PRELIMINARY;	PRT; 666 AA.
ID	Q7PZMS	
AC	Q7PZMS:	
RA	AAQ56703;	
DT	02-MAR-2004 (TREMBREL. 27, Created)	
DT	02-MAR-2004 (TREMBREL. 27, Last sequence update)	
DT	02-MAR-2004 (TREMBREL. 27, Last annotation update)	
DE	UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41).	
DE	Drosophila melanogaster (Fruit fly).	
OS	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles; Drosophila melanogaster (Fruit fly).	
OX	NCBI_TaxId=180454;	
RN	[1] TAXID=7227;	
RP	SEQUENCE FROM N.A.	
RA	PubMed=12829714;	
RA	RA: Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.; RT: Functional Characterization and Expression Analysis of Members of the UDP-GalNAc:Polypeptide N-Acetyl-galactosaminyltransferase Family from RT: Drosophila melanogaster. J. Biol. Chem. 278:35039-35048 (2003).	
DR	EMBL: AY266067; AAQ56703.1.	
DR	Glycosyltransferase; transferase.	
DR	SEQUENCE 666 AA; 7698 MW; 1D18362EE0UDBD196 CRC64;	
Query Match		
Best Local Similarity	49.0%;	Score 1360; DB 2; Length 666;
Matches	277;	Conservative 76; Mismatches 160; Indels 52; Gaps 15;
QY	67	DSQKTKQHDKEARRDAQRVNGEQRGPYPMTDQDAVR--ENGFTNTVSKTSL 123
Db	121	DASVKQWHDYTFEMEKDAKRVGLGEGGKASTLDDESQDLEKRMISLENGNALLSDISV 180
QY	124	NRSLPLDIRHPCNCNSKRYETLPNTSIIIPFHNGGWSLRLTVASVLSNNSPPELVAVIVLW 183
Db	181	NRSVPDIRHPLCRKKEVYKLPVTSVIIIFVNEYLSVLMRVSISLINSPELMKEILV 240
QY	184	DDPSDREHLKKPLDMM-LFPRSVTRLTTRKREGIRTRMGLGASVATGIVTIDSICEA 242
Db	241	DHSDREVLGELEYTYAEBFKAVRVLPRTEGLIGARAGARANATAEVLFILDSHVERA 300
QY	243	NYVNLWPLPLDRTARNRKTIVCPMDVHDFFRTETQDAMGAEDEMYKRPITPPE 302
Db	301	NYVNLWPLPLIEPIALNKRTAVCPFDIDVQDITNFHYAQ-DEGARGAFDWFPPYKRLPLPE 359
QY	303	LQKADPSPFESVMMAGGLPAVRKWFHLLGGIDPGLSEIWGQYELSPKVMQGGRMED 362
Db	360	DLK-HADPKSISIMAGGLFALSREFFLGELGYDEGLIWGGQYELSPKWMQGGRMED 418
QY	363	ICRSRVRGHYR---KTVYKPYGAVGSLARNLRVAYTVMDEAYEYIYORRER-YRHLIS 416
Db	419	APCSRIGIYGRGRNHQSPRK---GDIYHKNYKRVRAEWMDKBYKLYXSHGGGLYESVD 475
QY	417	AGDVAVQKLRSINCKSPKWMKTIKANDLPKYPVPPVPAAMGEIRVNGT-GLCAGT- 474
Db	476	PGDLTEQKAIKTRKLNCKSPKWMKVEAVADLMLKTPVPPDPSVYANGALQVNGNLCLOTL 535
QY	475	--KHGALG-----SPLRLGGCVRGREGGEAAWNMQVFTTWRDTRPGDQHTKRP 522
RESULT 10		
Q7PZMS	PRELIMINARY;	PRT; 599 AA.
ID	Q7PZMS	
AC	Q7PZMS:	
RA	Anopheles Genome Sequencing Consortium;	
RA	Submitted (MAR-2002) to the EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is	
CC	CC	
DR	EMBL; AAAB01008986; BAM00319.1; -	
DR	InterPro; IPR001773; Glyco_trans_2.	
DR	InterPro; IPR000772; Ricin_B_lectin.	
DR	Pfam; PF00535; Glycos_transf_2.1.	
DR	PFam; PF00652; Ricin_B_lectin; 3.	
DR	PROSITE; PS05231; RICIN_B_LECTIN; 1.	
FT	NON_TER 1	
SQ	SEQUENCE 599 AA; 69224 MW; DA646C182B143028 CRC64;	
Query Match		
Best Local Similarity	49.0%;	Score 1341.5; DB 2; Length 599;
Matches	257;	Conservative 104; Mismatches 175; Indels 25; Gaps 13;
QY	52	QGQHSRQKTTFLGQKUKDKEATRRAQDQRVNGEQRGPYPMTDQAV--DQAY 108
Db	46	GEGTGYAMPNV---AGEKIDWNYELIEESEKRGTKGEGHGRYKLUSEQDIALAKP 100
QY	109	RENGFNIYVSDKISLNRSLPDIRHPCNCNSKRYETLPNTSIIIPFHNGGWSLRLTVHSV 168
Db	101	KENGSYAVVSDMIALNRSPDIFRHPSCRMKEVYLKLPVTSVIIIFYNEHWSALLRTVSV 160
QY	169	LNRSPPELVAVIVLVDPSDRERHKKPLDM---ALFSPVRLTRKREGIRTRMLGAS 226
Db	161	LNRSPPELVAVIVLVDPSDRERHKKPLDM---ALFSPVRLTRKREGIRTRMLGAS 220
QY	227	VATGIVTFLDHSCEANVWMLWPLPLDRTARNRKTIVCPMDVHDFFRTETQDAMRG 286
Db	221	BARGDVLVLTDSHTEVNTWMLWPLPLDRTARDYTCVCPFDVHDFFRTETQDAMRG 279
QY	287	AFDNEYKRP---TPELQKADSPDPESSPVMAGGLFAVDRKWFWELSGYDPGLB1WGG 345
Db	280	AFDPKPYYKRLPLPGLD---DPTPKENSPVMAAGGLFAISAKFFWELGGDDEGIDWGG 337
QY	346	QELSPKVMQGGRMEDICRCSRVRGHYRCKVYKPYKPGAGS-LARNLKRVABVWMDAYEY 404
Db	338	QYERSFKWQGGGLVLDACSRVGHYKPYKPGAGS-LARNLKRVABVWMDAYEY 397
QY	405	YQRPPEYRHLASDADVOKKRSLSNCKSPKWMKTIKANDLPKYPVPPVPAAMGEIRVNGT 578
Db	577	CDVQIWDANAPYPTWMLWDCHSQGNOQWYDVRHKQLKGTEGRROLELLPPSQEAVNKC 636

RESULT 11

Q7DRO PRELIMINARY; PRT; 1003 AA.

ID Q7DRO; AC Q7DRO; DT 01-MAR-2004 (Tremblel. 26, Last sequence update)

DT 01-MAR-2004 (Tremblel. 26, Last annotation update)

DE AGC10655 (fragment)

GN Name=acgg47419; ORFNames=ENSANGG00000013497; OS Anopheles gambiae str. PEST. OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles. NCBI\_TaxID=180454; RN [1] RPPSEQUENCE FROM N.A.

RC STRAIN=PEST; RA Anopheles Genome Sequencing Consortium; RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

CC -I- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.

DR EMBL; AAB00108849; RA07231; 1; - DR InterPro; IPR00173; Glyco\_transf\_2. DR InterPro; IPR000772; Ricin B lectin. DR Pfam; PF00535; Glycos\_transf\_2. DR Pfam; PF00652; Ricin\_B lectin; 4. DR PROSITE; PS50231; RICIN\_B\_LRCTIN; 2. FT NON\_TER 1 115923 MW; 753EA50F567A4E13 CRC64; SQ SEQUENCE 1003 AA; [2]

Query Match Best Local Similarity 38.0%; Score 1247; DB 2; Length 1003; Matches 258; Conservative 77; Mismatches 179; Indels 28; Gaps 14; QY 73 DWIDKEATRRAQDQVGNGFQGRQGRPMPTDAERVD---QAYRENGFNIYVSDKISLNRLP 128 DR 27 DRYANVEQIQLDILRNGPSEQGKPTLSPEATSLRKELYKQFGNALLSDKISNSIA 86 QY 129 DIRHPPNCNSKRYLTPLNTSIIIPFHNBGWSLARTVHSVLSNSPPPELVAYTIVLDDSD 188 DR 87 DLRHPSCKSKSYRSHLPIASVWPPYEEHWSTLRTTIVSVLNNSPPHLIKEIILIVDGGT 146 DR 189 REIJKKPLDEMDAL-FPSVRLRIMKREGIIRTMGLASGVATGVDITFLDSHEANVWL 247 DR 147 KERFLNKLDYVQKLPKLVQKOPERGLIKARLAGAKTASGIVLITLSDHTEAGYWL 206 QY 248 PPIJDLTARNKTCITVCPMDIVDHDDFRVEIQADAMRGAFDHMMYKRIIPPELQAD 307 DR 207 PPIPLAENPKTCVCPDIDQTFOHPO-DEGGKLPDFTPHYKRVLINE-DRIS 264 QY 308 PSDPFPESPVMAGGLPFAVDRKFWLIGGDPGLETWGGROYELPEPKVMCGGRHMDICSR 367 DR 265 PTEPPFPSPVMAGGLPFAIGADFWLGGIDDELDIWGAEOYEISPKIOWGGRMLDAPCSR 324 DR 368 VGIYRKVRYPKVPGVS-LARNLKRVAEVMMRBYAAYYQRPPE-YHLSADIVVQKK 425 DR 325 FGHYRTVSPFPNSRKYDFTTRHHRVABWIMDYEKQVYDROPERVAKTDAGDMSKMT 384 QY 581 SLSLQQWLFHNTSTVLEKEN 601 DR 572 DKAQSQKWWGQYVWILQWID 592

RESULT 12

GLT4\_DROME STANDARD; PRT; 659 AA.

ID GLT4\_DROME AC Q8I42; Q8IQ11; DT 01-OCT-2004 (Rel. 45, Last sequence update)

DT 01-OCT-2004 (Rel. 45, Last annotation update)

DE N-Acetylgalactosaminyltransferase 4 (EC 2.4.1.1) (Protein-UDP acetylgalactosaminyltransferase 4) (UDP-GalNAc:polypeptide N-Acetylgalactosaminyltransferase 4) (pp-GalNAc 4).

GN Name=pgant4; ORFNames=CG31556; OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; RC Ephydriidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227; RN [1] RPPSEQUENCE FROM N.A., ENZYME ACTIVITY, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.

RC STRAIN=canon-S; TISSUE=Embryo; MEDLINE=22841110; Published=1829714; DOI=10.1074/jbc.M303836200; RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.; RA Brandon R.C., Rogers Y.H.-C., Blazej R.G., Charape M., Pfeiffer B.D., RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G., RA Abril J.F., Agbavani A., An H.-J., Andrews-Pfankoch C., Baldwin D., RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshkov S., RA Borkova D., Botcham M.R., Bouck J., Brokskein P., Brottier P., RA Burtis D.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., RA Cherry J.M., Cawley S., Daniske C., Davenport L.B., Davies P., RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Ditzt S.M., RA Dodson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., RA Durbin K.J., Evangelista C.C., Ferrera S., Ferreira C., Fleischmann W., RA Fosler C., Gabrialian A.E., Garg N.S., Geibart W.M., Glaser K., RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J., RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., RA Jallal M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., RA Laeko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X., RA Liu X., Mattei B., Mcintosh T.C., McLeod M.P., McPherson D., RA Merkulov G., Milisina N.V., Moberly C., Morris J., Moskneff A., RA Mount S.M., Moy M., Murphy B., Murphy L.M., Muzny D.M., Nelson D.L., RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M., RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirkas R., Tector C., Turner R., Venteo E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weintraub G.M., Weissbach J.,  
 RA Wodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*,"  
 RL *Science* 287:2185-2195 (2000).  
 RL [3]

RP REVISIONS.

RX MEDLINE:22426069; PubMed=12537572;

RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecsky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Berman B.P.,  
 RA Battencourt B.R., Celikiner S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richer J., Russo S., Schroeder A.J., Shu S.O.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 systematic review,"  
 RL *Genome Biol.* 3:RESEARCH0083.1-RESEARCH0083.22 (2002).  
 RL [4]

RN SEQUENCE OF 16-659 FROM N.A.

RX PubMed=11925150; DOI=10.1074/jbc.M202684200;

RA Schwienek T., Bennett E.P., Flores C., Thacker J., Hollmann M.,  
 RA Reis C.A., Behrens J.J., Mandel U., Keck L., Schaefer M.A.,  
 RA Haelemaen K., Zubarev R., Reepstorff P., Burchell J.M.,  
 RA Taylor-Papadimitriou J., Hollingsworth M.A., Clausen H.,  
 RT "Functional conservation of subfamilies of putative UDP-N-  
 acetylglactosamine:polypeptide N-acetylgalactosaminyltransferases in  
 RT *Drosophila*, *Caenorhabditis elegans* and mammals. One subfamily  
 RT composed of 1(2)35Ba is essential in *Drosophila*,"  
 RL *J. Biol. Chem.* 277:22623-22638 (2002).

CC --I- FUNCTION: Glycopeptide transferase involved in O-linked  
 CC Oligosaccharide biosynthesis, which catalyzes the transfer of an  
 CC N-acetyl-D-galactosamine residue to an already glycosylated  
 peptide. In contrast to other proteins of the family, it does not  
 act as a peptide transferase that transfers GalNAc onto serine or  
 threonine residue on the protein receptor, but instead requires  
 the prior addition of a GalNAc on a peptide before adding  
 additional GalNAc moieties. Some peptide transferase activity is  
 however not excluded, considering that its appropriate peptide  
 substrate may remain unidentified. Prefers the diglycosylated  
 CC Muc5AC-3/13 as substrate.

CC --I- COFACTOR: Manganese and calcium (By similarity).

CC --I- PATHWAY: Glycosylation.

CC --I- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
 CC similarity).

CC --I- TISSUE SPECIFICITY: Expressed in developing oocytes and egg  
 chambers.

CC --I- DEVELOPMENTAL STAGE: Expressed during embryonic, larval, pupal and  
 adult stages. Weakly expressed during early embryonic stages but  
 displays a dramatic increase at 12-24 h of embryonic development.  
 Continues to be in adult but displays much lower levels in the  
 female adult as compared with the male.

CC --I- DOMAIN: There are two conserved domains in the glycosyltransferase  
 region: the N-terminal domain (domain A, also called GM1 motif),  
 which is probably involved in manganese coordination and substrate  
 binding and the C-terminal domain (domain B, also called  
 Gal/GalNAc-T motif), which is probably involved in catalytic  
 reaction and UDP-Gal binding (By similarity).

CC --I- DOMAIN: The ricin type lectin domain binds to GalNAc and  
 contributes to the glycopeptide specificity (By similarity).  
 CC --I- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
 CC subfamily.

CC --I- SIMILARITY: Contains 1 ricin B-type lectin domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).

CC

CC

DR EMBL; AR003579; RAN10370.1; -;

DR EMBL; AR324732; RAN17551.1; -;

DR FlyBase; PBSP0051956; pgant4;

DR InterPro; IPR00173; Glyco\_trans\_2.

DR InterPro; IPR00897; RicinB-like.

DR InterPro; IPR000772; RicinB\_lectin.

DR Pfam; PF00535; Glycos\_transf\_2\_1.

DR Pfam; PF0652; Ricin\_B\_lectin\_3.

DR PROSITE; PS50231; RICIN\_B\_LBCTIN\_1.

KW Calcium; Glycosyltransferase; Golgi stack; Lectin; Manganese;

KW Signal-anchor; Transferase; Transmembrane.

FT DOMAIN 1 25

FT TRANSMEM 26 48

FT DOMAIN 49 659

FT DOMAIN 192 303

FT DOMAIN 360 422

FT DOMAIN 511 644

FT DISULFID 524 541

FT DISULFID 571 588

FT DISULFID 615 632

FT CARBOHYD 172 172

FT CARBOHYD 194 194

FT CARBOHYD 544 544

FT CARBOHYD 580 580

FT CARBOHYD 647 647

FT CARBOHYD 659 AA: 75805 MW: 0182D42REBA1DD2 CRC64;

FT SEQUENCE

Query Match 36.2%; Score 1186; DB 1; Length 659;  
 Best Local Similarity 41.2%; Pred. No. 2.98-87; Gaps 21;  
 Matches 266; Conservative 106; Mismatches 216; Indels 58; Gaps 21;

QY 2 RKEKEKLLQAVVALVAAVLVLPNGLWLR-----EROPDG--TPGGSGAAVA 48  
 QY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 20 KRYVKGKLUKVKVLLWVTVLWVTVLWVTERMMNAELTEQLDPNGLPITPVRANITH 79  
 QY 49 PA---AGGSHSRQKETPFLPDQGKL-----KWDKKAIRRAQRVNGEODR 94  
 QY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 80 PTRKAPRPPQDRNQVVDIPRSQDKLQGLFRLPEPKGERKWDVYAAMEDRKRSGFGERQV 139  
 QY 95 PYPM--TDAERVD-DAVRENGFNYVSK1SLRSLPDTPHNCNSKQRLTETLPNTSII 151  
 QY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 140 AVKLNPDKEQLEKHYEKGNGFLISDRISVRSVPLDRLEACKTRKYLAKUNIVF 199  
 QY 152 PFHNGWGSLLRTHSVLNRSPPVLEAVLVLVDDPSDRHLLKELEDVIAL-FPS-VRL 209  
 QY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 200 IFFNEHFRNTLRSIYSVNRTPPELQVILVLDGSEMDVNLQQLDPPVQHQHFLVTRV 259  
 QY 210 RTKKEKGLRTRMIGASVAVGDTVTPLDHCEANWNLPLDRAARNRKTRIVCPMDVI 269  
 QY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 260 RNPERRQLIGARIAGAKAVAVQGVNFFPSDHSIEVNVWPLPPLIEAINPKISTCPMDVI 319  
 QY 270 DHDPRYERQAGDAMGAFDEMWTYKR1PPELQOKAEDSPDPEPSPWMMGLPFLVDRKVR 329  
 QY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 320 SHEDPSVFSGNKGDKGARGFDPWMLQYKQLPVLPE-DALDKSMYRSPVWMMGLFAINTDP 378  
 QY 330 WBLGGDPGLIGEAGBQYETTSFKWVMMGGRMEDIPCSRYCHIYR-KVIVYKVPGVS-LA 387  
 QY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 379 WDLGSDYDQDINGESEQVELSKWVMMGLDLPVCPSCRAHIFRGPMKQGRGNPRHGFIV 438  
 QY 388 RNLKRVAEVWDEYVYTORPE-YRHSAGDVAVOKLRSIANSCKSPRMWFLAWDL 446  
 QY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 439 KKHKRVAEVWDEYKQYVKGDPKTYDNLADGLTRQGRVRLKCKSFHWFMTTEVA 498  
 QY 447 PKFYPVPEPPAAMGEETRNGTGL-CAPTRKGALGSPPLLEGCVTRGRGGRGAAMNNQVTF 505  
 QY :|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 499 LVKFPVPEPPAAMGEETRNGTGL-CAPTRKGALGSPPLLEGCVTRGRGGRGAAMNNQVTF 556



RP REVISIONS.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Mirz S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hredcky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B., Celik N., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richer J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.,  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.", J:RESEARCH0083, 1-RESEARCH0083, 22 (2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley; TISSUE=Embryo;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RL Genome Biol. 3:RESEARCH0080, 1-RESEARCH0080, 8 (2002).  
 CC Stapleton M., Carlson J.W., Brokstein P., Yu C., Chame M., Park S., Wan K.H.,  
 RA George R.A., Guarin H., Kromann B., Pacieb J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celtnier S.E.,  
 RA "A *Drosophila* full-length cDNA resource.", J:RESEARCH0080, 1-RESEARCH0080, 8 (2002).  
 CC -!- FUNCTION: May catalyze the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor (By similarity).  
 CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -!- COFACTOR: Manganese and calcium (By similarity).  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR\_LOCALIZATION: Type II membrane protein. Golgi (By  
 CC similarity).  
 CC -!- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region; the N-terminal domain (domain A, also called Gm1 motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC Gal/GalNAc-T motif), which is probably involved in catalytic  
 CC reaction and UDP-Gal binding (By similarity).  
 CC -!- DOMAIN: The ricin B-type lectin domain binds to GalNAc and  
 CC contributes to the glycopptide specificity (By similarity).  
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
 CC subfamily.  
 CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no  
 CC way modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; AB003806; AAF57964.2; -.  
 DR HSSP; P26514; IKNM.  
 DR FLYBase; FBgn050463; CG30463.  
 DR InterPro; IPR01173; Glyco\_trans\_2.  
 DR InterPro; IPR00901; Putat\_T\_DNA\_bind.  
 DR InterPro; IPR008997; Ricin\_B\_lectin.  
 DR Pfam; PF0055; Glycos\_transf\_2\_1.  
 DR Pfam; PF00652; Ricin\_B\_lectin\_3.  
 DR SMART; SM00458; Ricin\_1.  
 DR ROSTE; PS50231; RICIN\_B\_LECTIN\_1.  
 KW Calcium; Glycocalyx; transferase; Transmembrane; Golgi stack; Lectin; Manganese;  
 KW Signal-anchor; Transmembrane; Lectin; Manganese;  
 FT DOMAIN 1 11  
 FT TRANSMEM 1 11  
 FT 12 31  
 FT  
 FT DOMAIN 32 650  
 FT DOMAIN 208 317  
 FT DOMAIN 378 440  
 FT DOMAIN 521 643  
 FT 43 90  
 RP  
 PT DISULFID 535 554  
 PT DISULFID 577 590  
 PT DISULFID 616 631  
 PT CARBOHYD 321 321  
 PT CARBOHYD 373 373  
 PT CONFLICT 145 145  
 PT CONFLICT 454 454  
 PT CONFLICT 472 472  
 PT SEQUENCE 650 AA; 73192 MN; AC847736ADIC07CA CRC64;  
 PT  
 Query Match 35.8%; Score 1172.5; DB 1; Length 650;  
 Matches 272; Conservative 40.2%; Prod. No. 3.6e-86; Indels 113; GapB 22;  
 QY 2 RKEKEKRLQGAVALVIAALAVLMLPNNGLWALYRERQDGTGGGSAVAPAGGSS---- 55  
 Db 6 RRRSTTIVKLVAFALAINFCIAFLVYTPTDTRRAAQCEGAGASAGSARPGVGGAGGLDP 65  
 QY 56 -----HSQKKTFL-----GDPQKUKDWD----- 76  
 Db 66 TALALRNRPAGEDFGINGNVIGGGQ-KQAHDBADTPPTVGKHKADIQAERMRKKAEO 123  
 QY 77 -KEAIRRDAQRV-----GNOBQGPY-----PMTD-----AERDQAYRENGENIYSDK 120  
 Db 124 PKKPKQBSKKVIDPPANPFEEENPGELGPVRLEKEMSPBEMKKAVDGSQTNAFQYQDSL 183  
 QY 121 ISLNRSLPDRTRHPNCNSK-RYLETLPLNTSIIIPHNNEGSLLRTVHSVNLNRSPPELVAE 179  
 Db 184 ISVRLTLPDPRDAKCDKARYLTLNPKDVIICFHNEAWTULLRTVHSVLDSPHEHLIGK 243  
 QY 180 IIVLVDPSIREHLLKPLPDEYALMAPPVSLRTRKREGJLRTMIGASATGVDITFLSH 239  
 Db 244 IILVDDSPDMPLHQLKQLEDFYAXPKVQVIRGOKREGJLIGANHAKSPVLYLISH 303  
 QY 240 CEAUNWNPFLDITARNRKTICPMPMVIDHDFPYS-TQAGDAMRGPARDWMMYKIP 298  
 Db 304 CECTEGWLEPLDITARNSTTIVCPVIVDVTISDETLLEYHVDRSGCUNVGFDWNLQFSWHIP 363  
 QY 299 IPEELQKADPS-DPFESPVMAGGLFADRKWVBLG3DGPGLB1WGBQYETSPKVMC 356  
 Db 364 VPERERKERNSTAAVPVSPTMAGGLFSDRFPDRGTYDSDPINDGENLESPKTM 423  
 QY 357 GGRMDPDRPSRVGHLYKRVYKPYKPGAS-LARNLKTRAEVNDYEAQYIYOR---RPE 411  
 Db 424 GGTBLIIVPCSHVGHIFRKPSPKYPRSGVNLKGNVSLRJWVHDEBYSQYHHRIGNDKGD 483  
 QY 412 YRHLISAGDVAVQVKLRSUNLNSKSKWPMKTIANDLPKTYPPVPEP-AAAWGRIRYCTG 469  
 Db 484 W----GVDYSDRDRKLNDLCKSPKWPYLDN----YPELPIPGDSVAHGBIRNLYG 531  
 QY 470 --LCDTKEG----ALGSPPLRLEGCVGRGRGEAANWMDMQVFTWREDIRPDPQHTRK 521  
 Db 532 GRTCLDAPAGKQHQKQKAVGT---YPCIRQGKQW----MLSKAGBIRADS---- 576  
 QY 522 PCPDA1SHSPVTLVDCSMKGKQWTKD-KTLYHGVSGSCMDCSSESDHRAFMNTCP 580  
 Db 577 -CLDYAG--KDVTLFGCHGCKGKQFWTTRNTQHLHGTSGKCLAISSSKDOKULMECSA 633  
 QY 581 SLSLQOWLREHTNSTVL 597  
 Db 634 SLSRQWLTENYDSSKL 650  
 RP  
 RESULT 15  
 Q7Q815 PRELIMINARY PRT; 645 AA.  
 ID Q7Q815  
 AC Q7Q815;  
 DT 01-MAR-2004 (TREMbl, 26, Created)  
 DT 01-MAR-2004 (TREMbl, 26, Last sequence update)  
 DT 01-MAR-2004 (TREMbl, 26, Last annotation update)  
 DE EbiP504 (Fragment)  
 GN Name=ebiP504; ORName=ENSANGG0000003900;  
 OS Anopheles gambiae str. PEST.

GN	Name=acG55772; ORFNames=ENSANG000000013981;
OC	Anopheles gambiae ST-363; PEST;
OC	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles
NCBI_TaxID=180454;	[1]
RN	SEQUENCE FROM N.A.
RP	STRAIN_PEST;
RA	Anopheles Genome Sequencing Consortium; STRAIN_PEST;
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC	-1- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.
CC	EMBL; AAAB01008944; EA010180.1; -.
DR	InterPro; IPR001173; Glyco trans 2.
DR	InterPro; IPR001172; Ricin B lectin.
DR	InterPro; IPR001172; Ricin B lectin.
DR	Pfam; PF00532; Glycos transf_2; 1.
DR	Pfam; PF00532; Ricin B lectin; 3.
DR	Pfam; PF00532; Ricin B lectin; 3.
DR	PROSITE; PSS0231; RICIN_B_LECTIN; 1.
FT	PROSITE; PSS0231; RICIN_B_LECTIN; 1.
SQ	SEQUENCE 645 AA; 73459 MW; A4DFDB6C816B51CO CRC64;
FT	NOV_TER 1 1
SQ	SEQUENCE 518 AA; 58964 MW; 00F8821266B395FE CRC64;
Query Match	35.4%; Score 1159; DB 2; Length 645;
Best Local Similarity	44.3%; Pred. No 4.4e-85;
Matches	251; Conservative 79; Mismatches 193; Indels 44; Gaps 19;
Qy	38 GTPCGSGAAVAPAAAGQGSHSRQKTPPLDGQKQKWDKEARRRDAQVRGN GEQGP- 95
Db	101 GDPGCGYGGG--GCGRDAADSSMPPRTY--RPOBLKWKWQAPV--AENYGRGPGEWGPV 151
Qy	96 -YPMTDAERDVQAVRRENGENIYVSDKISLNSRSLDIRHNCNSKRYLETLPNTSI1.PFH 154
Db	152 KIPANQOELMKERKEKERNQNLASDMLNRLSTDVRHDKCKKHPAKLPTSVIVFH 211
Qy	155 NEGMWSLARTVHSYLNRSRPPLEVAEVLYDFFSDREHKKPLDYMALFP-SVRLRKK 213
Db	212 NEAWSTLRTIWTIWSVNRSPRPLKETILVTDASBRERHLGRQLEBYVRUTLPVPTVLRGK 271
Qy	214 REGLRTMIGASVATGDTVIFLDSCHEANVNLPLDRIANRKTIVCPMDVHD 273
Db	272 RSGLRARULGAKHKGKQVITFUDAHCECTEGMILEPLALARLIVDRKTVCPIDVSDT 331
Qy	274 FRYETQAGDAMRGIAFDWEMYKRP1.PIPP-ELOKA--DPSDPFSPVMAGGLFAVDKWFW 330
Db	332 FEVT-ASDPTWGGFNWKUNFRMVRVPAAREMQRNHDRTAPLTPTMAGGLFSDIDRVDY 390
Qy	331 ELLGGDPGLEIWMGQEQYEISPKVWMMGGMEDIPICSRVGHYIKVVKPVKPGV--LAR 388
Db	391 EIGSYDBGNDIWMGSENLENPSFVWCGGLEIAPCSRVGHVFRKSTPVSFGCTSTQIN 450
Qy	389 NLKRAEWWDEAYVYIYKORRPEYRHLASAGDVAVQKQKLSSLNCKSFKNMFTKIAWDLPK 448
Db	451 NNARLAEWLDGWNSEFPYNNINGGARKASAGDVSRRALRERLCKSKFRWLENI---- 504
Qy	44.9 FYPVPEPRAA--GEIRVGTGCAADPRTKRGAGSPPLREGCYRGRGERAAMNNMQVFP 506
Db	505 -YEVSQMPDYYFGLGIRVKTINCLDTWGRKESNEKICSSY-C-HGGLG----GNOVRAFT 557
Qy	507 WREDIRPGDPOHKTCKFCDAISHTSPVTLYDCHSMKGNLWY-RKDLYLHVGSCMD 565
Db	558 KRIHQIMSDDN----CLOASNALGPGVNLVRCHGMGGNQEWIVTDEEKTIKHNSGNLT 611
Qy	566 -CSBSDHHR1-FMNTCCNPSSLTQOMLFB 590
Db	612 RASEDDPSTPLRPCNYSB-GQOWLMQ 637
RESULT 16	
GLUT_HUMAN	
ID	GLUT_HUMAN STANDARD; PRT; 518 AA.
07Q048	PRELIMINARY; PRT; 518 AA.
AC	Q0048;
DT	01-MAR-2004 (TREMBLrel. 26, Created)
DT	01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE	AGCP9928 (Fragment).
RESULT 17	
GLUT_HUMAN	
ID	GLUT_HUMAN STANDARD; PRT; 559 AA.
AC	Q10472; Q867U7; Q9UMB6;
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DE	Polypeptide N-acetylglucosaminyltransferase 1 (EC 2.4.1.41)
DE	(Protein-UDP acetylglucosaminyltransferase 1) (UDP-
DE	GalNAcC-polypeptide N-acetylglucosaminyltransferase 1) (Polypeptide

DB GalNAc transferase 1) (GalNAc-T1) (pp-Gantase 1) [Contains: DE Polypeptide N-acetylglactosaminyltransferase 1, soluble form].  
 RN [1]  
 RN SEQUENCE FROM N.A. (ISOFORM 1).  
 RN TISSUE=salivary gland;  
 RX MEDLINE=96115928; PubMed=8607019;  
 RA Meijer J.A., Naylor J.M., Baker C.A., Thomsen D.R., Homa F.L.,  
 RA Elhammar A.P.;  
 RT "cDNA cloning, expression, and chromosomal localization of a human  
 RT UDP-GalNAc:polypeptide, N-acetylglactosaminyltransferase.;"  
 RL J. Biochem. 118:568-574 (1995).  
 RN  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.  
 RX MEDLINE=96028800; PubMed=7522619;  
 RA White T., Bennett E.P., Takio K., Soerensen T., Bonding N.,  
 RA Clausen H.;  
 RT "Purification and cDNA cloning of a human Ump-N-acetyl- $\alpha$ -D-galactosaminyltransferase.;"  
 RT galactosaminyltransferase.;"  
 RL J. Biol. Chem. 270:2416-24165 (1995).  
 RN  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=ENS; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Krausser R.D., Collins S.P., Wagner L., Shekman C.M., Schuler G.D.,  
 RA Aitschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.P., Jordan H., Moore T., Max S.J., Wang J., Hsieh F.,  
 RA Diatchenko L., Matsubara K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Tsaihoku S., Carninci P., Frangie C.,  
 RA Raha S.S., Iqbalian N.A., Peters G.J., Abramson R.D., Millahy S.J.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fainey J., Heitton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shchepetko Y., Bouffard G.G.,  
 RA Brakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimmold J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krywinski M.J., Skalska U., Smalls D.E.,  
 RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.;"  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [4]  
 RP SEQUENCE OF 512-559 FROM N.A. (ISOFORM 1).  
 RX MEDLINE=96285863; PubMed=8727794;  
 RA Meijer J.A., Drong R.F., Homa F.L., Slichtom J.L., Elhammar A.P.;  
 RT "Organization of human UDP-GalNAc:polypeptide, N-acetylglactosaminyltransferase gene and a related processed  
 RT pseudogene.;"  
 RL Glycopatology 6:231-241 (1996).  
 RN [5]  
 RP FUNCTION.  
 RX Pubmed=9295285;  
 RA Wandal H.H., Hassan H., Mirgorodskaya E., Kristensen A.K.,  
 RA Roopstorff P., Bennett E.P., Nielsen P.A., Hollingsworth M.A.,  
 RA Burchell J., Taylor-Papadimitriou J., Clausen H.;  
 RT "Substrate specificities of three members of the human UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylglactosaminyltransferase family: GalNAc-T1, -T2, and -T3.;"  
 RL J. Biol. Chem. 272:23503-23514 (1997).  
 RN [6]  
 RP SUBCELLULAR LOCATION.  
 RP PubMed=9394011;  
 RA Roottger S., White J., Wandal H.H., Olivo J.-C., Stark A.,  
 RA Bennett E.P., Whitehouse C., Berger E.G., Clausen H., Nilsson T.;  
 RT "Localization of three human polypeptide GalNAc-transferases in HeLa  
 cells suggests initiation of O-linked glycosylation throughout the

RT Golgi apparatus.;"  
 RL J. Cell Sci. 111:5-60 (1998).  
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked oligosaccharide biosynthesis, the transfer of an N-acetyl-D-galactosamine residue to a serine or threonine residue on the protein receptor. Has a broad spectrum of substrates for peptides such as EA2, Muc5AC, Muc1a, Mucin and Muc7.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide = UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium.  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi; resides evenly across the Golgi stack. A secreted form also exists.  
 CC -1- ALTERNATIVE PRODUCTS: Event=Alternative splicing; Named isoforms=2;  
 CC Name=1; IsoID=Q10472-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoID=Q10472-2; Sequence=SP 011200;  
 CC Note=No experimental confirmation available;  
 CC TISSUE SPECIFICITY: Widely expressed. Expressed in all tissues tested.  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase region: the N-terminal domain (domain A, also called GTI motif), which is probably involved in manganese coordination and substrate binding and the C-terminal domain (domain B, also called Gal/GalNAc-T motif), which is probably involved in catalytic reaction and UDP-Gal binding (By similarity).  
 CC -1- DOMAIN: The ricin B-type lectin domain directs the glycopeptide specificity. It is required in the glycopeptide specificity of substrates, suggesting that it triggers the catalytic domain to act on GalNAc-glycopeptide substrates (By similarity).  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T subfamily.  
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi-sib.ch/announce/> or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
 CC  
 DR EMBL; U01514; AAC50327.1;  
 DR EMBL; U01518; CAM59380.1;  
 DR EMBL; U01519; AAC50328.1;  
 DR EMBL; U01520; AAC50329.1;  
 DR EMBL; U01521; AAC50330.1;  
 DR EMBL; U01522; AAC50331.1;  
 DR EMBL; U01523; AAC50332.1;  
 DR EMBL; U01524; AAC50333.1;  
 DR EMBL; U01525; AAC50334.1;  
 DR EMBL; U01526; AAC50335.1;  
 DR EMBL; U01527; AAC50336.1;  
 DR EMBL; U01528; AAC50337.1;  
 DR EMBL; U01529; AAC50338.1;  
 DR EMBL; U01530; AAC50339.1;  
 DR EMBL; U01531; AAC50340.1;  
 DR EMBL; U01532; AAC50341.1;  
 DR EMBL; U01533; AAC50342.1;  
 DR EMBL; U01534; AAC50343.1;  
 DR EMBL; U01535; AAC50344.1;  
 DR EMBL; U01536; AAC50345.1;  
 DR EMBL; U01537; AAC50346.1;  
 DR EMBL; U01538; AAC50347.1;  
 DR EMBL; U01539; AAC50348.1;  
 DR EMBL; U01540; AAC50349.1;  
 DR EMBL; U01541; AAC50350.1;  
 DR EMBL; U01542; AAC50351.1;  
 DR EMBL; U01543; AAC50352.1;  
 DR EMBL; U01544; AAC50353.1;  
 DR EMBL; U01545; AAC50354.1;  
 DR EMBL; U01546; AAC50355.1;  
 DR EMBL; U01547; AAC50356.1;  
 DR EMBL; U01548; AAC50357.1;  
 DR EMBL; U01549; AAC50358.1;  
 DR EMBL; U01550; AAC50359.1;  
 DR EMBL; U01551; AAC50360.1;  
 DR EMBL; U01552; AAC50361.1;  
 DR EMBL; U01553; AAC50362.1;  
 DR EMBL; U01554; AAC50363.1;  
 DR EMBL; U01555; AAC50364.1;  
 DR EMBL; U01556; AAC50365.1;  
 DR EMBL; U01557; AAC50366.1;  
 DR EMBL; U01558; AAC50367.1;  
 DR EMBL; U01559; AAC50368.1;  
 DR EMBL; U01560; AAC50369.1;  
 DR EMBL; U01561; AAC50370.1;  
 DR EMBL; U01562; AAC50371.1;  
 DR EMBL; U01563; AAC50372.1;  
 DR EMBL; U01564; AAC50373.1;  
 DR EMBL; U01565; AAC50374.1;  
 DR EMBL; U01566; AAC50375.1;  
 DR EMBL; U01567; AAC50376.1;  
 DR EMBL; U01568; AAC50377.1;  
 DR EMBL; U01569; AAC50378.1;  
 DR EMBL; U01570; AAC50379.1;  
 DR EMBL; U01571; AAC50380.1;  
 DR EMBL; U01572; AAC50381.1;  
 DR EMBL; U01573; AAC50382.1;  
 DR EMBL; U01574; AAC50383.1;  
 DR EMBL; U01575; AAC50384.1;  
 DR EMBL; U01576; AAC50385.1;  
 DR EMBL; U01577; AAC50386.1;  
 DR EMBL; U01578; AAC50387.1;  
 DR EMBL; U01579; AAC50388.1;  
 DR EMBL; U01580; AAC50389.1;  
 DR EMBL; U01581; AAC50390.1;  
 DR EMBL; U01582; AAC50391.1;  
 DR EMBL; U01583; AAC50392.1;  
 DR EMBL; U01584; AAC50393.1;  
 DR EMBL; U01585; AAC50394.1;  
 DR EMBL; U01586; AAC50395.1;  
 DR EMBL; U01587; AAC50396.1;  
 DR EMBL; U01588; AAC50397.1;  
 DR EMBL; U01589; AAC50398.1;  
 DR EMBL; U01590; AAC50399.1;  
 DR EMBL; U01591; AAC50400.1;  
 DR EMBL; U01592; AAC50401.1;  
 DR EMBL; U01593; AAC50402.1;  
 DR EMBL; U01594; AAC50403.1;  
 DR EMBL; U01595; AAC50404.1;  
 DR EMBL; U01596; AAC50405.1;  
 DR EMBL; U01597; AAC50406.1;  
 DR EMBL; U01598; AAC50407.1;  
 DR EMBL; U01599; AAC50408.1;  
 DR EMBL; U01600; AAC50409.1;  
 DR EMBL; U01601; AAC50410.1;  
 DR EMBL; U01602; AAC50411.1;  
 DR EMBL; U01603; AAC50412.1;  
 DR EMBL; U01604; AAC50413.1;  
 DR EMBL; U01605; AAC50414.1;  
 DR EMBL; U01606; AAC50415.1;  
 DR EMBL; U01607; AAC50416.1;  
 DR EMBL; U01608; AAC50417.1;  
 DR EMBL; U01609; AAC50418.1;  
 DR EMBL; U01610; AAC50419.1;  
 DR EMBL; U01611; AAC50420.1;  
 DR EMBL; U01612; AAC50421.1;  
 DR EMBL; U01613; AAC50422.1;  
 DR EMBL; U01614; AAC50423.1;  
 DR EMBL; U01615; AAC50424.1;  
 DR EMBL; U01616; AAC50425.1;  
 DR EMBL; U01617; AAC50426.1;  
 DR EMBL; U01618; AAC50427.1;  
 DR EMBL; U01619; AAC50428.1;  
 DR EMBL; U01620; AAC50429.1;  
 DR EMBL; U01621; AAC50430.1;  
 DR EMBL; U01622; AAC50431.1;  
 DR EMBL; U01623; AAC50432.1;  
 DR EMBL; U01624; AAC50433.1;  
 DR EMBL; U01625; AAC50434.1;  
 DR EMBL; U01626; AAC50435.1;  
 DR EMBL; U01627; AAC50436.1;  
 DR EMBL; U01628; AAC50437.1;  
 DR EMBL; U01629; AAC50438.1;  
 DR EMBL; U01630; AAC50439.1;  
 DR EMBL; U01631; AAC50440.1;  
 DR EMBL; U01632; AAC50441.1;  
 DR EMBL; U01633; AAC50442.1;  
 DR EMBL; U01634; AAC50443.1;  
 DR EMBL; U01635; AAC50444.1;  
 DR EMBL; U01636; AAC50445.1;  
 DR EMBL; U01637; AAC50446.1;  
 DR EMBL; U01638; AAC50447.1;  
 DR EMBL; U01639; AAC50448.1;  
 DR EMBL; U01640; AAC50449.1;  
 DR EMBL; U01641; AAC50450.1;  
 DR EMBL; U01642; AAC50451.1;  
 DR EMBL; U01643; AAC50452.1;  
 DR EMBL; U01644; AAC50453.1;  
 DR EMBL; U01645; AAC50454.1;  
 DR EMBL; U01646; AAC50455.1;  
 DR EMBL; U01647; AAC50456.1;  
 DR EMBL; U01648; AAC50457.1;  
 DR EMBL; U01649; AAC50458.1;  
 DR EMBL; U01650; AAC50459.1;  
 DR EMBL; U01651; AAC50460.1;  
 DR EMBL; U01652; AAC50461.1;  
 DR EMBL; U01653; AAC50462.1;  
 DR EMBL; U01654; AAC50463.1;  
 DR EMBL; U01655; AAC50464.1;  
 DR EMBL; U01656; AAC50465.1;  
 DR EMBL; U01657; AAC50466.1;  
 DR EMBL; U01658; AAC50467.1;  
 DR EMBL; U01659; AAC50468.1;  
 DR EMBL; U01660; AAC50469.1;  
 DR EMBL; U01661; AAC50470.1;  
 DR EMBL; U01662; AAC50471.1;  
 DR EMBL; U01663; AAC50472.1;  
 DR EMBL; U01664; AAC50473.1;  
 DR EMBL; U01665; AAC50474.1;  
 DR EMBL; U01666; AAC50475.1;  
 DR EMBL; U01667; AAC50476.1;  
 DR EMBL; U01668; AAC50477.1;  
 DR EMBL; U01669; AAC50478.1;  
 DR EMBL; U01670; AAC50479.1;  
 DR EMBL; U01671; AAC50480.1;  
 DR EMBL; U01672; AAC50481.1;  
 DR EMBL; U01673; AAC50482.1;  
 DR EMBL; U01674; AAC50483.1;  
 DR EMBL; U01675; AAC50484.1;  
 DR EMBL; U01676; AAC50485.1;  
 DR EMBL; U01677; AAC50486.1;  
 DR EMBL; U01678; AAC50487.1;  
 DR EMBL; U01679; AAC50488.1;  
 DR EMBL; U01680; AAC50489.1;  
 DR EMBL; U01681; AAC50490.1;  
 DR EMBL; U01682; AAC50491.1;  
 DR EMBL; U01683; AAC50492.1;  
 DR EMBL; U01684; AAC50493.1;  
 DR EMBL; U01685; AAC50494.1;  
 DR EMBL; U01686; AAC50495.1;  
 DR EMBL; U01687; AAC50496.1;  
 DR EMBL; U01688; AAC50497.1;  
 DR EMBL; U01689; AAC50498.1;  
 DR EMBL; U01690; AAC50499.1;  
 DR EMBL; U01691; AAC50500.1;  
 DR EMBL; U01692; AAC50501.1;  
 DR EMBL; U01693; AAC50502.1;  
 DR EMBL; U01694; AAC50503.1;  
 DR EMBL; U01695; AAC50504.1;  
 DR EMBL; U01696; AAC50505.1;  
 DR EMBL; U01697; AAC50506.1;  
 DR EMBL; U01698; AAC50507.1;  
 DR EMBL; U01699; AAC50508.1;  
 DR EMBL; U01700; AAC50509.1;  
 DR EMBL; U01701; AAC50510.1;  
 DR EMBL; U01702; AAC50511.1;  
 DR EMBL; U01703; AAC50512.1;  
 DR EMBL; U01704; AAC50513.1;  
 DR EMBL; U01705; AAC50514.1;  
 DR EMBL; U01706; AAC50515.1;  
 DR EMBL; U01707; AAC50516.1;  
 DR EMBL; U01708; AAC50517.1;  
 DR EMBL; U01709; AAC50518.1;  
 DR EMBL; U01710; AAC50519.1;  
 DR EMBL; U01711; AAC50520.1;  
 DR EMBL; U01712; AAC50521.1;  
 DR EMBL; U01713; AAC50522.1;  
 DR EMBL; U01714; AAC50523.1;  
 DR EMBL; U01715; AAC50524.1;  
 DR EMBL; U01716; AAC50525.1;  
 DR EMBL; U01717; AAC50526.1;  
 DR EMBL; U01718; AAC50527.1;  
 DR EMBL; U01719; AAC50528.1;  
 DR EMBL; U01720; AAC50529.1;  
 DR EMBL; U01721; AAC50530.1;  
 DR EMBL; U01722; AAC50531.1;  
 DR EMBL; U01723; AAC50532.1;  
 DR EMBL; U01724; AAC50533.1;  
 DR EMBL; U01725; AAC50534.1;  
 DR EMBL; U01726; AAC50535.1;  
 DR EMBL; U01727; AAC50536.1;  
 DR EMBL; U01728; AAC50537.1;  
 DR EMBL; U01729; AAC50538.1;  
 DR EMBL; U01730; AAC50539.1;  
 DR EMBL; U01731; AAC50540.1;  
 DR EMBL; U01732; AAC50541.1;  
 DR EMBL; U01733; AAC50542.1;  
 DR EMBL; U01734; AAC50543.1;  
 DR EMBL; U01735; AAC50544.1;  
 DR EMBL; U01736; AAC50545.1;  
 DR EMBL; U01737; AAC50546.1;  
 DR EMBL; U01738; AAC50547.1;  
 DR EMBL; U01739; AAC50548.1;  
 DR EMBL; U01740; AAC50549.1;  
 DR EMBL; U01741; AAC50550.1;  
 DR EMBL; U01742; AAC50551.1;  
 DR EMBL; U01743; AAC50552.1;  
 DR EMBL; U01744; AAC50553.1;  
 DR EMBL; U01745; AAC50554.1;  
 DR EMBL; U01746; AAC50555.1;  
 DR EMBL; U01747; AAC50556.1;  
 DR EMBL; U01748; AAC50557.1;  
 DR EMBL; U01749; AAC50558.1;  
 DR EMBL; U01750; AAC50559.1;  
 DR EMBL; U01751; AAC50560.1;  
 DR EMBL; U01752; AAC50561.1;  
 DR EMBL; U01753; AAC50562.1;  
 DR EMBL; U01754; AAC50563.1;  
 DR EMBL; U01755; AAC50564.1;  
 DR EMBL; U01756; AAC50565.1;  
 DR EMBL; U01757; AAC50566.1;  
 DR EMBL; U01758; AAC50567.1;  
 DR EMBL; U01759; AAC50568.1;  
 DR EMBL; U01760; AAC50569.1;  
 DR EMBL; U01761; AAC50570.1;  
 DR EMBL; U01762; AAC50571.1;  
 DR EMBL; U01763; AAC50572.1;  
 DR EMBL; U01764; AAC50573.1;  
 DR EMBL; U01765; AAC50574.1;  
 DR EMBL; U01766; AAC50575.1;  
 DR EMBL; U01767; AAC50576.1;  
 DR EMBL; U01768; AAC50577.1;  
 DR EMBL; U01769; AAC50578.1;  
 DR EMBL; U01770; AAC50579.1;  
 DR EMBL; U01771; AAC50580.1;  
 DR EMBL; U01772; AAC50581.1;  
 DR EMBL; U01773; AAC50582.1;  
 DR EMBL; U01774; AAC50583.1;  
 DR EMBL; U01775; AAC50584.1;  
 DR EMBL; U01776; AAC50585.1;  
 DR EMBL; U01777; AAC50586.1;  
 DR EMBL; U01778; AAC50587.1;  
 DR EMBL; U01779; AAC50588.1;  
 DR EMBL; U01780; AAC50589.1;  
 DR EMBL; U01781; AAC50590.1;  
 DR EMBL; U01782; AAC50591.1;  
 DR EMBL; U01783; AAC50592.1;  
 DR EMBL; U01784; AAC50593.1;  
 DR EMBL; U01785; AAC50594.1;  
 DR EMBL; U01786; AAC50595.1;  
 DR EMBL; U01787; AAC50596.1;  
 DR EMBL; U01788; AAC50597.1;  
 DR EMBL; U01789; AAC50598.1;  
 DR EMBL; U01790; AAC50599.1;  
 DR EMBL; U01791; AAC50600.1;  
 DR EMBL; U01792; AAC50601.1;  
 DR EMBL; U01793; AAC50602.1;  
 DR EMBL; U01794; AAC50603.1;  
 DR EMBL; U01795; AAC50604.1;  
 DR EMBL; U01796; AAC50605.1;  
 DR EMBL; U01797; AAC50606.1;  
 DR EMBL; U01798; AAC50607.1;  
 DR EMBL; U01799; AAC50608.1;  
 DR EMBL; U01800; AAC50609.1;  
 DR EMBL; U01801; AAC50610.1;  
 DR EMBL; U01802; AAC50611.1;  
 DR EMBL; U01803; AAC50612.1;  
 DR EMBL; U01804; AAC50613.1;  
 DR EMBL; U01805; AAC50614.1;  
 DR EMBL; U01806; AAC50615.1;  
 DR EMBL; U01807; AAC50616.1;  
 DR EMBL; U01808; AAC50617.1;  
 DR EMBL; U01809; AAC50618.1;  
 DR EMBL; U01810; AAC50619.1;  
 DR EMBL; U01811; AAC50620.1;  
 DR EMBL; U01812; AAC50621.1;  
 DR EMBL; U01813; AAC50622.1;  
 DR EMBL; U01814; AAC50623.1;  
 DR EMBL; U01815; AAC50624.1;  
 DR EMBL; U01816; AAC50625.1;  
 DR EMBL; U01817; AAC50626.1;  
 DR EMBL; U01818; AAC50627.1;  
 DR EMBL; U01819; AAC50628.1;  
 DR EMBL; U01820; AAC50629.1;  
 DR EMBL; U01821; AAC50630.1;  
 DR EMBL; U01822; AAC50631.1;  
 DR EMBL; U01823; AAC50632.1;  
 DR EMBL; U01824; AAC50633.1;  
 DR EMBL; U01825; AAC50634.1;  
 DR EMBL; U01826; AAC50635.1;  
 DR EMBL; U01827; AAC50636.1;  
 DR EMBL; U01828; AAC50637.1;  
 DR EMBL; U01829; AAC50638.1;  
 DR EMBL; U01830; AAC50639.1;  
 DR EMBL; U01831; AAC50640.1;  
 DR EMBL; U01832; AAC50641.1;  
 DR EMBL; U01833; AAC50642.1;  
 DR EMBL; U01834; AAC50643.1;  
 DR EMBL; U01835; AAC50644.1;  
 DR EMBL; U01836; AAC50645.1;  
 DR EMBL; U01837; AAC50646.1;  
 DR EMBL; U01838; AAC50647.1;  
 DR EMBL; U01839; AAC50648.1;  
 DR EMBL; U01840; AAC50649.1;  
 DR EMBL; U01841; AAC50650.1;  
 DR EMBL; U01842; AAC50651.1;  
 DR EMBL; U01843; AAC50652.1;  
 DR EMBL; U01844; AAC50653.1;  
 DR EMBL; U01845; AAC50654.1;  
 DR EMBL; U01846; AAC50655.1;  
 DR EMBL; U01847; AAC50656.1;  
 DR EMBL; U01848; AAC50657.1;  
 DR EMBL; U01849; AAC50658.1;  
 DR EMBL; U01850; AAC50659.1;  
 DR EMBL; U01851; AAC50660.1;  
 DR EMBL; U01852; AAC50661.1;  
 DR EMBL; U01853; AAC50662.1;  
 DR EMBL; U01854; AAC50663.1;  
 DR EMBL; U01855; AAC50664.1;  
 DR EMBL; U01856; AAC50665.1;  
 DR EMBL; U01857; AAC50666.1;  
 DR EMBL; U01858; AAC50667.1;  
 DR EMBL; U01859; AAC50668.1;  
 DR EMBL; U01860; AAC50669.1;  
 DR EMBL; U01861; AAC50670.1;  
 DR EMBL; U01862; AAC50671.1;  
 DR EMBL; U01863; AAC50672.1;  
 DR EMBL; U01864; AAC50673.1;  
 DR EMBL; U01865; AAC50674.1;  
 DR EMBL; U01866; AAC50675.1;  
 DR EMBL; U01867; AAC50676.1;  
 DR EMBL; U01868; AAC50677.1;  
 DR EMBL; U01869; AAC50678.1;  
 DR EMBL; U01870; AAC50679.1;  
 DR EMBL; U01871; AAC50680.1;  
 DR EMBL; U01872; AAC50681.1;  
 DR EMBL; U01873; AAC50682.1;  
 DR EMBL; U01874; AAC50683.1;  
 DR EMBL; U01875; AAC50684.1;  
 DR EMBL; U01876; AAC50685.1;  
 DR EMBL; U01877; AAC50686.1;  
 DR EMBL; U01878; AAC50687.1;  
 DR EMBL; U01879; AAC50688.1;  
 DR EMBL; U01880; AAC50689.1;  
 DR EMBL; U01881; AAC50690.1;  
 DR EMBL; U01882; AAC50691.1;  
 DR EMBL; U01883; AAC50692.1;  
 DR EMBL; U01884; AAC50693.1;  
 DR EMBL; U01885; AAC50694.1;  
 DR EMBL; U01886; AAC50695.1;  
 DR EMBL; U01887; AAC50696.1;  
 DR EMBL; U01888; AAC50697.1;  
 DR EMBL; U01889; AAC50698.1;  
 DR EMBL; U01890; AAC50699.1;  
 DR EMBL; U01891; AAC50700.1;  
 DR EMBL; U01892; AAC50701.1;  
 DR EMBL; U01893; AAC50702.1;  
 DR EMBL

FT DOMAIN	29	559	Luminal, catalytic (Potential).
FT DOMAIN	115	225	catalytic subdomain A.
FT DOMAIN	287	347	catalytic subdomain B.
FT DOMAIN	429	551	Ricin B-type lectin.
FT SITE	141	141	Not glycosylated (By similarity).
FT DISULFID	442	459	By similarity.
FT DISULFID	482	497	By similarity.
FT DISULFID	523	540	By similarity.
FT CARBOHYD	95	117	N-linked (GlcNAc. . .) (By similarity).
FT CARBOHYD	117	117	O-linked (Potential).
FT CARBOHYD	118	118	O-linked (Potential).
FT CARBOHYD	119	119	O-linked (Potential).
FT CARBOHYD	288	288	O-linked (Potential).
FT CARBOHYD	541	541	N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD	552	552	N-linked (GlcNAc. . .) (By similarity).
FT VARSPPLIC	106	559	Missing (in isoform 2).
FT			/PfamID:SP 011200.
FT SEQUENCE	559	AA:	64219 MW: CD68118@B201E5B CRG64;
Query Match			Score: 1125; DB: 1; Length: 559;
Best Local Similarity			34.03%; Score: 1125; DB: 1; Length: 559;
Matches			232; Conservative: 87; Mismatches: 170; Indels: 38; Gaps: 16;
Qy	88	GlNEQGRP--YPTDPAERDVQAYRENGENIYSDKISLNRSIPLRIPNCNSKRYLETP	145
Db	57	GPQEMGKPVWVPIKEDQEKKEMPKINQFNLMASMAEMLANRSLPDRVLGCKTKVYPDNLP	116
Qy	146	NTSIIIPPHNEGKSSLLRTVHSYLANSRSPPELVAELVLUVDDPSDRHKKPLDYM-ALPP	204
Db	117	TRSVVPHVNEAVSLLTARTVHSYINRSPHMEELIVLVLDASERDFLKRPLSSYVKKV	176
Qy	205	SYRILRKKEGGLRTTRIMGASVATGVDITFLDSHCRANVNLPLPLDRIARRKTVCP	264
Db	177	PVHVRMEQRSGLIRAKLKGAVSKGQVITFLDAHCECTVGMLEPLLARIKDRRTVCP	236
Qy	265	MIDVTDIDPFRYETQAG-DAMKGAFDMEMYKRIPIP--ELQKAPSDPESPPWAGG	320
Db	237	IIDVSDDTPEY--MASSDMTYXGGFNNWLNFRPTVPRGDRDTLPRTPTWAGG	294
Qy	321	LFAVDRKWFELGGYDPLGIEWGQYKISPKWMMCGRMEDIPCSRIGHVIRKVKPVK	380
Db	295	LFSIDRKYFOEIGTYDAGMDINGBNLEISFRIWQCGTTLBIVTCSHVGHVPRKATPTP	354
Qy	381	PAGVS--LARNLKVRAEVYDFAEYTORRPRYHRASAGDVAVOKURRSSLACKSKF	438
Db	355	PGGTGQIINKNNERLRLAEVWMDDEFKNNFFYIISPGVTKVYDIDSSRVLRLHKLUQPKPSW	414
Qy	439	MTKIAWD--LPKTYPPTEPPAAGWEIRVNGTGLCPTKKGALGSPLEGCVRGREGAA	496
Db	415	LENIYPPDQSPQIPRY------FSIGETRNVETNQCLDMARKENEKYGIFNC-HGMG--	463
Qy	497	WNNMQVFTFTWREDIIRPDPDQHPTKKFCDAISHTSPVTLYDCHSMKGNQWLWKVKKDK-TL	555
Db	464	-GNQVFSYANKEIRTD-----LCLDVSKLNGPVTMLKCHLKGQNLWEDPVKLT	515
Qy	556	YHVGSGCND-CSESDRRI-FMNTCNPESLTLQWLFETNSTLKEF	600
Db	516	QHVNINQCLDQKATEEDSQQPSIRDQN-GSRSQQWLLR--NVTULPEI	559
RESULT 18			
GT1_PIG			
GT1_PIG			
STANDARD;			PRT: 559 AA.
ID			
Q29121;			
16-OCT-2001 (Rel. 40, Created)			
DT			
16-OCT-2001 (Rel. 40, Last sequence update)			
DT			
01-OCT-2004 (Rel. 45, Last annotation update)			
DE			
Polypeptide N-acetylgalactosaminyltransferase 1 (EC 2.4.1.41)			
(Protein-UDP-N-acetylglucosaminyltransferase 1) (UDP-			
GlcNAc-polypeptide N-acetylgalactosaminyltransferase 1) (Polypeptide			
GalNAc transferase 1) (GalNAc-T1) (PP-GalNase 1) [Contains: Polypeptide N-acetylgalactosaminyltransferase 1, soluble form].			
Name=GalNT1;			
FT CHAIN	41	559	
FT			
FT			
FT			
FT DOMAIN	1	8	Acetylglucosaminyltransferase 1, soluble form (By similarity).
FT TRANSMEM	9	28	Cytoplasmic (Potential).
FT			
FT DOMAIN	29	559	Signal-anchor for type II membrane protein (Potential).
FT DOMAIN	115	225	Luminal, catalytic (Potential).
FT DOMAIN	285	347	Catalytic subdomain A.
FT DOMAIN	429	551	Ricin B-type lectin.
FT SITE	141	141	Not glycosylated (Probable).
FT DISULFID	442	459	By similarity.
FT DISULFID	482	497	By similarity.

FT	DISU8FID	523	540	By similarity.
FT	CARBODY	95	95	N-linked (GlcNAc-.) (By similarity).
FT	CARBODY	117	117	O-linked (Potential).
FT	CARBODY	118	118	O-linked (Potential).
FT	CARBODY	119	119	O-linked (Potential).
FT	CARBODY	288	288	O-linked (Potential).
FT	CARBODY	552	552	N-linked (GlcNAc-.) (By similarity).
SQ	SEQUENCE	559 AA;	64118 MW;	68C2D261A516846 CRC64;
Query Match		34.1;	Score 117;	DB 1; Length 559;
Best Local Matches		43.6%;	Pred. No. 9. 4e-82;	90; Mismatches 16; Indels 38; Gaps 16;
88	GNGBQSRP--YMTDARVDDQYRQENPENIYSDKSILARSFDPDRHNCNSKRYETLP	145		
57	GPQGEMGKPKVPIKEDQDKOMKEMKINGENPLMASEMILANRSPDVRLEGKTKVYDNP	116		
146	NTSIIIPPFNHBGSLLRTVHSYLNRSRPPLEYTIVLVDDESDREHLLKKPLDYM-ALFP	204		
Qy	117	TSVWVTFHNEAWSTLRTVHSYLNRSRPPLEYTIVLVDDESDREHLLKKPLDYM-ALFP	176	
Qy	205	SVRILRKTKREGIIRTMGLASVATGIVTFLDSHCEANWMLPPIIDRARNKTVCP	264	
Qy	177	EVHIVRQEORSLGILARLKGAVASQVITFIDAHCECTVGWLLEPLIARIKDRKTVCP	236	
Db	265	MIDVIRHDDFRVETQAG-DAMGAFDMEYVYKIPITP--BLQKADPSDFESPAGG	320	
Db	321	LFAVDKRNFWELGGYDGELETWGEYEBISFKWMCGRMEDIPCRSERGHTYKRYPKV	380	
Qy	295	LPSIDRDFQELGTYDAGMDIWGEGNLBISFRIWQCGTLEITVCHVGFVKRATPYTP	354	
Db	381	PAQVS-LARNLKVRSVYMDSYAEVYVQRRRBYRHSAGDVAVQKLLRSSLACKSKEKWP	438	
Qy	355	PGGTGQIINKNRRLRQAVWMDERKTFVYIISGVTQYDGGISRGLRHKQCRPSWY	414	
Db	439	MTRKAMW-LPKRYPPVEPPAAGWGERLNGVGLADTKHGAQGSPRLLEGVRGEEA	496	
Qy	415	LENIYPPDSQIIPRYH----SSLGEIRANVETQCDLMARKENKEKGIFNC-HGMG--	463	
Db	497	WNMNOVPTFTEDRIPGDPOHTKFFDAISHTSPYTLYDCHSMKGNOVQYKOK-TL	555	
Qy	464	--GNQVFSYANKERITD----LCUDVSKLNGVPMILKCHLKKNOLWEDVPKUTL	515	
Db	556	YHPVSSCMD-CSESHRI-FNATCNISLQTQWLFHENTNSVLEKF	600	
Qy	516	OHVNSQGLDQKATEQSDQVSPRDCS-GSRSQWLR--NVLPEIF	559	
RESULT 19				
GLTL-RAT				
ID	GLTL-RAT		STANDARD:	PRT;
AC	Q1073;			559 AA.
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-2004 (Rel. 45, Last annotation update)			
DB	Polypeptide N-acetylglucosaminyltransferase 1 (EC 2.4.1.41) (Protein-UDP acetylglucosaminyltransferase 1) (UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase 1) (Polypeptide GalNAc transferase 1) (GalNAc-T1) (pp-GalTase 1) [contains: Name=GalNT1; DE Polypeptide N-acetylglucosaminyltransferase 1, soluble form].			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxId:10116;			
[1]	SEQUENCE FROM N.A.			
RP	STRAINV-Sprague-Dawley; TISSUE=Sublingual gland;			
RP	MEDLINE=96318029; PubMed=374818;			
RA	Hagen F., Gregorie C.A., Tabak L.A.; "Cloning and sequence homology of a rat UDP-GalNAc:polypeptide N-			
RT				
RT	acetylglucosaminyltransferase.;"			
RL	Glycoconj. J. 12:901-909 (1995).			
RL	[2]			
RP	MUTAGENESIS OF CYS-106; CYS-212; CYS-214; CYS-235; CYS-330; CYS-339 AND CYS-400.			
RP	PUBMED=12139709;			
RA	Tenne M., Toba S., Kezdy F.J., Elhammar A.P., Kurosaka A.; Teno M., Kezdy F.J., Elhammar A.P., Kurosaka A.;			
RA	PubMed=1249318;			
RA	PMID:1249318;			
RT	"Identification of two cysteine residues involved in the binding of UDP-GalNAc to UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase polypeptide with multiple acceptor sites.";			
RL	J. Biol. Chem. 277:47088-47096 (2002).			
RN	[4]			
RP	MUTAGENESIS OF ASP-442; CYS-497; CYS-523; CYS-540; ASP-444; GLY-455; PHE-468; ASN-466; ASP-464 AND ASP-525.			
RP	PubMed=1234335; DOI=10.1074/jbc.M027369200;			
RA	Tenne M., Saeki A., Kezdy F.J., Elhammar A.P., Kurosaka A.; Teno M., Kezdy F.J., Elhammar A.P., Kurosaka A.;			
RT	"The lectin domain of UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase 1 is involved in O-glycosylation of a polypeptide with multiple acceptor sites.";			
RL	PMID:1249318;			
RA	Teno M., Kezdy F.J., Elhammar A.P., Kurosaka A.; Teno M., Kezdy F.J., Elhammar A.P., Kurosaka A.;			
RT	"Function of the lectin domain of polypeptide N-acetylglucosaminyltransferase 1.";			
RT	Biochem. Biophys. Res. Commun. 298:755-759 (2002).			
CC	-I- FUNCTION: Catalyses the initial reaction in O-linked oligosaccharide biosynthesis, the transfer of an N-acetyl-D-galactosamine residue to a serine or threonine residue on the protein receptor. Has a broad spectrum of substrates for peptides such as EA2, MucA, Mucia, Mucib and Mu7.			
CC	-I- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide = UDP + N-acetyl-D-galactosaminyl-polypeptide.			
CC	-I- COFACTOR: Manganese and calcium.			
CC	-I- PATHWAY: Glycosylation.			
CC	-I- SUBCELLULAR LOCATION: Type II membrane protein, Golgi; resides evenly across the Golgi stack. A secreted form also exists (By similarity).			
CC	-I- TISSUE SPECIFICITY: Heart, brain, spleen, liver, skeletal muscle and kidney.			
CC	-I- DOMAIN: There are two conserved domains in the glycosyltransferase region: the N-terminal domain (domain A, also called G1 motif), which is probably involved in manganese coordination and substrate binding and the C-terminal domain (domain B, also called Gai/GalNAc-T motif), which is probably involved in catalytic reaction and UDP-Gal binding.			
CC	-I- DOMAIN: The ricin B-type lectin domain directs the glycopeptide specificity. It is required in the glycopeptide specificity of enzyme activity but not for activity with naked peptide substrates, suggesting that it triggers the catalytic domain to act on GalNAc-glycopeptide substrates.			
CC	-I- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T CC			
CC	-I- SIMILARITY: Contains 1 ricin B-type lectin domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce">http://www.isb-sib.ch/announce</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL; U35890; AAC52511.1; -.			
DR	EMBL; U35890; AAC52511.1; -.			
DR	EMBL; U35890; AAC52511.1; -.			
DR	InterPro; IPR001173; Glyco_trans_2.			
DR	InterPro; IPR00999; Ricin_B_lectin.			
DR	InterPro; IPR00772; Ricin_B_lectin.			
DR	Pfam; PF00335; Glycos_transf_2; 1.			
DR	Pfam; PF0052; Ricin_B_lectin; 3.			
DR	SMART; SM00458; RICIN; 1.			



CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -!- COFACTOR: Manganese and calcium.  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi; resides  
 CC evenly across the Golgi stack. A secreted form also exists (By  
 CC similarity).  
 CC -!- TISSUE SPECIFICITY: Colistrium contains a soluble form.  
 CC -!- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called *Gn1* motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC *Gn1/Gnac-T* motif), which is probably involved in catalytic  
 CC reaction and UDP-Gal binding (By similarity).  
 CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.  
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 2. *Gn1ac-T*  
 CC subfamily.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; Li7437; AAA30532.1; -.  
 DR PIR; A45987; A45987; -.  
 DR InterPro; IPR00173; Glyco\_trans\_2..  
 DR InterPro; IPR00772; Ricin\_B\_lectin; 3.  
 DR SMART; SM00458; RICIN\_N; 1.  
 DR PROSITE; PS50231; RICIN\_B\_LCITIN; 1.  
 DR Calcium; Direct protein sequencing; Glycoprotein; Glycosyltransferase;  
 KW Golgi stack; Lectin; Manganese; Signal-anchor; Transferase;  
 KW Transmembrane. 1 40 Removed in soluble polypeptide N-  
 PROPEP  
 FT CHAIN 41 559 Acetylgalactosaminyltransferase, 1,  
 FT DOMAIN 1 8 Catalytic subdomain A.  
 FT DOMAIN 9 28 Cyttoplasmic (potential) signal-anchor for type II membrane  
 TRANSMEM  
 FT DOMAIN 29 559 luminal (potential).  
 FT DOMAIN 115 225 Catalytic subdomain A.  
 FT DOMAIN 285 347 Catalytic subdomain B.  
 FT SITE 141 141 Not glycosylated (Probable).  
 FT DISULPID 442 459 By similarity.  
 FT DISULPID 482 497 By similarity.  
 FT CARBOHYD 523 540 N-linked (GlcNAc, -) (Probable).  
 FT CARBOHYD 141 141 O-linked (Potential).  
 FT CARBOHYD 117 117 O-linked (Potential).  
 FT CARBOHYD 118 118 O-linked (Potential).  
 FT CARBOHYD 119 119 O-linked (Potential).  
 FT CARBOHYD 288 288 O-linked (Potential).  
 FT CARBOHYD 552 552 N-linked (GlcNAc, -) (Probable).  
 FT MUTAGEN 95 95 N->Q: Induces decrease in glycosylation.  
 FT MUTAGEN 125 125 H->A: Induces a strong decrease in  
 activity.  
 FT MUTAGEN 137 137 H->A: No effect.  
 FT MUTAGEN 141 141 N->Q: No effect.  
 FT MUTAGEN 146 146 H->A: No effect.  
 FT MUTAGEN 179 179 H->A: No effect.  
 FT MUTAGEN 211 211 H->A: Loss of function.

FT MUTAGEN 228 228 H->A: No effect.  
 FT MUTAGEN 341 341 H->A: Induces a strong decrease in  
 activity.  
 FT MUTAGEN 344 344 H->A: Loss of function.  
 FT MUTAGEN 404 404 H->A: No effect.  
 FT MUTAGEN 427 427 H->A: No effect.  
 FT MUTAGEN 460 460 H->A: No effect.  
 FT MUTAGEN 498 498 H->A: No effect.  
 FT MUTAGEN 499 499 H->A: No effect.  
 FT MUTAGEN 517 517 H->A: No effect.  
 FT MUTAGEN 552 552 N->Q: Induces decrease in glycosylation.  
 FT SEQUENCE 559 AA; 64192 MW; 83853IC4DE69B40 CRC64;  
 Qy Query Match 34.0%; Score 1115; DB 1; Length 559;  
 Qy Best Local Similarity 43.6%; Pred. No. 1.4e-81; Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;  
 Db Db 57 GPGENGKPKVPIKPSQEDQEKMKFQINQFQKISLNRSLPDIRHNCNSKYLETLIP 145  
 Qy 88 GNGNGQGR--YPMTDAAEYDQAYQRENGNIVYQDKISLNRSLPDIRHNCNSKYLETLIP 145  
 Qy 146 NTSTIIPFNEGNGWASLRTVHSVANRSPPELIVBIVLUDSFREHKLKPDLDM-ALFP 204  
 Qy 147 TTSVWVIFNEANESTLRTVHSVNRSPPELIVBIVLUDSFREHKLKPDLIPVPTMAGG 320  
 Db 117 TTSVWVIFNEANESTLRTVHSVNRSPPELIVBIVLUDSFREHKLKPDLIPVPTMAGG 294  
 Qy 205 SVRLRTKREGREGIRTRMAGASVATGVDITFLDSHCEANVNPPLRIARAKTIVCP 264  
 Qy 231 LFAVDRKWFELGGYDPLGLBWIWGCGBOYELSPKVMWCGGMDTIPCSRQHGTYRKV 380  
 Db 177 PVVHVRMSEGRSGLIRARLKGAAVSKQGVITFLDHCETCTVGWLEPLLARIKHDRTVCP 236  
 Qy 265 MIDVHDUDFRYETQAG-DAMRGAFDWMEMYKRIPIP--ELQADDSDFPESPVMMAGG 320  
 Db 237 IIVDSDPFEY--MAGSMTYGGFWKUNFRMVPVQPREMDRKGDTLPVPTMAGG 294  
 Qy 381 PAGVS--LARNLKAVALVWDEAYVYQRRPEYRHLASGDDVAVOKURSSLNCKSFKPWF 438  
 Db 355 PGGTQQTINKNRRLAENVDEKFKNPFFVITSPGTVKQYGDISRLGHLKQCRPFMSW 414  
 Qy 439 MTKLAWD--LPKFPPVPEPPAAMGEIRVNGTGLCADYKHALGSPPLLEGCVRGRESA 496  
 Db 415 LENITYPSDQIPHY-----FSGIRAVNETVQCDNMARKENKEVQGIFNCG-- 463  
 Qy 497 WNNMQVFTFWREDIRPQGPQHNTKKFCPDAISHTSPVUTYDCHSMKGHQWLKVRKDK-TL 555  
 Db 464 --GNQFVSTANKERITDD-----LCIDVSKLNGPVWLMKCHLKGQQLWEIDPVKCTL 515  
 Qy 556 YHPVSGCMD-CSSESDHRI-FMNTCNPSLTLQNLFEHNTVLEKF 600  
 Db 516 QHVNNSQCLDKATBDQVPSIRDCS-SSRSQQQLR--NVTLPEIF 559  
 RESULT 21  
 GTI\_MOUSE ID\_GTI\_MOUSE STANDARD; PRT: 559 AA.  
 AC 008912; Q7TND1;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; last sequence update)  
 DT 01-OCT-2004 (Rel. 45; last annotation update)  
 DE Polypeptide N-acetylgalactosaminyltransferase 1 (EC 2.4.1.41)  
 DE Polypeptide UDP-acetylgalactosaminyltransferase 1 (UDP-  
 DE GlcNAc:polypeptide N-acetylgalactosaminyltransferase 1) (Polypeptide  
 DE GlcNAc transferase 1) (Gn1ac-T1) (Ppp-GaNAcT1) [Contains:  
 DE Polypeptide N-acetylgalactosaminyltransferase 1, soluble form].  
 DE Name=Gn1ac-T1;  
 GN Mus musculus (Mouse).  
 OC Buxtorfia; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TAXID=10090;  
 RN [1]  
 SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.

RP





RT and mouse cDNA sequences.";  
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 R.L. [2]  
 R.N.  
 R.P.  
 R.C.  
 TISSUE=Eye;  
 MEDLINE=22341132; PubMed=12454917;  
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 Richardson P., *"Genetic and genomic tools for Xenopus research: The NIH Xenopus*  
*initiative."* Dev. Dyn. 225:384-391(2002).  
 RL [3].  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Eye;  
 RA Klein S., Strausberg R.;  
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC074224; AAH74234.1; -  
 DR InterPro; IPR001173; Glyco\_trans\_2.  
 DR InterPro; IPR008997; Ricin\_B lectin.  
 DR Pfam; PF0055; Glycos\_transf\_2, 1.  
 DR Pfam; PF00652; Ricin\_B Lectin; 3.  
 DR SMART; SM00438; RICIN\_1;  
 DR PROSITE; PS50231; RICIN\_B\_LACTIN\_1.  
 KW Hypothetical protein.  
 SEQUENCE 556 AA; 63908 MW; 3A9FEE4A39AB655B CRC64;

Query Match 33.9%; Score 112.5; DB 2; Length 556;  
 Best Local Similarity 42.8%; Pred. No. 2.2e-81; Mismatches 181; Indels 41; Gaps 16;  
 Matches 234; Conservative 91; Mismatches 181; Indels 41; Gaps 16;

QY 64 FLGDCOKLWDKE--AIRRDAQR--VNGEGRGP--YPMFTAERYDQAYBENGTY 116  
 DB 27 YFSBCNCKCDKDRSLLPALARAVISRSFPGPGLGKAVITPKDQEKMKELFKINQFLM 86

QY 117 VSDKISLNSLDPDRHPIGNSKXLEYDTNTSIIIPFNEGMSLLRTVHSVANSPRL 176  
 DB 87 ASDJALNLNSLDPDRLEGECKTKVYDPLNTSIVFHNEAWSTLRTVHSVINSPPRL 146

QY 177 VAETVLVDDFSDRLEHLKPELDYMA-LFFPSVRLRTKRGEGLIRTRMMGASVATGDIVTP 235  
 DB 147 ISEIILVDDASERPLRKPFLNPLKPELDYMA-LFFPSVRLRTKRGEGLIRTRMMGASVATGDIVTP 206

QY 236 LDSHEANTNWLPLFLDLRARNRKTIVPMDVHDHPFRTYQAG-TAMRGARDWMTY 294  
 DB 207 LDAAHCECTFGWLBPLALARIKEDRKTIVCPIDVSDTPEY-MAGSMTYGGENWKLNF 264

QY 295 KRITPP--BLOQADSPFESVMAGLFLAVDRKRMWELGGDPLGKINGSQYESP 351  
 DB 265 RWYVPPDREMRDRKEDRTRPVRPTMAGLSDKVKYRFLGTVDSGMDIWGHNLEMSF 324

QY 352 KVWMCGGRMEDIPCSRVGMYRKYVPUKVYPAGVS--LARNLKRVAEVMDYBAYIYOR 409  
 DB 325 RIWQCGSLEAVTCSHNVAVPKATPTPFGTCHVWIKNRQLAEVMDPDRDPFVPS 384

QY 410 PEYRHLASDADYVOKLKSINCKSFKWMKIAWDLKFKYPPVPEPPA--ANGBIRNG 467  
 DB 385 PGVVKYDGDVSEKALRNLKCNKPNFSWLYETV-----YPSQIIPRYSFLGEIRNVE 437

QY 468 TGLGADTKRKGALGSPRLRLEGVGRGEAMANNOVPTIWTREDIRPGRQHTKCFPAI 527  
 DB 438 TNQCLDNNMKERNKRNKGKTFNC-HGMG----GNQVFSTADKSRITDD-----LCLDVS 485

QY 528 SHTSPTVLTDCHSKMGKQWVKYDKD-TLXHPVSGSCDCSESDRHIT--MNTCPSSIT 584  
 DB 486 RLMPVPMKCHHNGQWVYDHLILRHINSQCDPADDKMPPTIKSCN-GRS 544

QY 585 QWLFEH 591  
 DB 545 QWLRN 551

RESULT 23

Q6GMS1 PRELIMINARY; PRT; 556 AA.

ID Q6GMS1  
 AC 06GMS1;  
 DT 05-JUL-2004 (TREMBLrel. 27, Created)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
 DE Hypothetical protein.

OS Xenopus laevis (African clawed frog).

OC Eutayota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 OC Xenopodinae; Xenopus.

RN [1] NCBI\_TaxID=8355;

RP SEQUENCE FROM N.A.

RC TISSUE=Eye;

RX MEDLINE=22388257; PubMed=1247932;

RA Strasburg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins P.S., Wagner L., Shemmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Matsunaga K., Farmer A.A., Robin G.M., Hong L.,  
 RA Stapleton M., Soakes M.B., Bonaldo M.F., Cabant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki T., Carminci P., Prange R.,  
 RA Rana S.S., Logueillo N.A., Peters G.J., Abramson R.D., Mulahy S.J.,  
 RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shverchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzwinowski M.I., Skalska U., Smailus D.E., Schnurch A., Schein J.E.,  
 RA Jones S.J., Marra M.A., Smailus D.E., Schnurch A., Schein J.E.,  
 RT "Generation and initial analysis of more than 15,000 full-length human

AAQ56702 PRELIMINARY, PRT; 617 AA.  
 ID ARQ56702; AC Q8C93; Q8BL4; Q8Y73;  
 AC AAQ56702; DT 02-MAR-2004 (TREMBLrel. 27, Created)  
 DT 02-MAR-2004 (TREMBLrel. 27, Last sequence update)  
 DT 02-MAR-2004 (TREMBLrel. 27, Last annotation update)  
 DE UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41).  
 DE 2.4.1.41).  
 DE PGANT5.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Mecoceridae;  
 OC Drosophilidae; Drosophila.  
 OC NCBITaxonID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX Published=2029714;  
 RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.;  
 RT "Functional Characterization and Expression Analysis of Members of the UDP-GalNAc:Polypeptide N-Acetylgalactosaminyltransferase Family from Drosophila melanogaster.",  
 RL J. Biol. Chem. 278:35039-35048 (2003).  
 DR EMBL: AK268066; ARQ56702.1; -.  
 KW Glycosyltransferase; Transferase.  
 SQ SEQUENCE 617 AA; 70482 MW; CADATAYA53B773E3 CRC64;  
 Query Match 33.9%; Score 1111.5; DB 2; Length 617;  
 Best Local Similarity 41.9%; Pred. No. 3e-81;  
 Matches 243; Conservative 82; Mismatches 194; Indels 61; Gaps 18;  
 Qy 33 EROPUGTPCGSGGAAYAPAGGGSHSRQKCFPFLGQK-----LKDMDHEKAIIRDAQ 85  
 Db 69 EKSLSLDEGG-----ALIMQG-----FASGGGISMTPSUVLKKPLAPSVOEAKG 114  
 Qy 86 RVGNGBQGRP-----YPTMDAERVDQAYRENGNNTIVYSDKISLRSIDPDIRPNCNSKRYLET 143  
 Db 115 K--PCEMGKVKIPADKMDLKEKEKQENQFLILLASDMSILNRLSDVRLRGCRKRYASK 172  
 Qy 144 LPNTSIIIPFNEGNSSSLRVTYHSVNLRSRPELVAEIVLVDDFSDREHLKKPLDEYMAF 203  
 Db 173 LPTTSIVIVHNEAWTLLKTVWSVINRSPRALIKEILVDDASERDFLQKQSBYVAKL 232  
 Qy 204 P-SVHLRTRKREGRLTRTMGLGASVATGDTYIETLSDHCEANVNMPLPLDRJARNRKTV 262  
 Db 233 PVKTFVRLTRKTRSLGIRARLIGAERVSGEVNTFLDAHCCTEGWLEPLKARRVQRTW 292  
 Qy 263 CPMIVDIDHDPFRYETOQADMGRGDWEMTYKRPPIPPE--LOKADSPDPFSPVMAG 319  
 Db 293 CPTIDVISDETFEYIT--ASDPTTWGCPNWKUNFRWTRVPSREMARUNNDTAPLRTPTMAG 351  
 Qy 320 GOLFADVRKWWIWEGLGGDGPGLIWEIWGCGEBOYESTSFKVNMCGGRMEDIPCSRVOHLYRKVVK 379  
 Db 352 GULFSIDKQFVEIGSYDEGMDIWGGNLEMFSRIVQCGGILEIPCSHVQHVFQDKSPY 411  
 Qy 380 VPGVS--LARNLKVEAVWMDVEAYKQRPEYRHLSSGDAVYOKKRSRSLACKSFK 437  
 Db 412 FPGGVAKIVLHNARVAEVNLEDEWRDFFYNSWSTGARKASAGDVSDRKALDRKCKSFR 471  
 Qy 438 FMTKIAWDLKFKYPRVERPAAM--GEIRNGTGCAPIHGAALGSPLEGCTGRGSEA 495  
 Db 472 YLENV-----VPSIMPLDYYVGLBIRNAETCLDTMGRKCKNEKEVSYC-HGLG-- 521  
 Qy 496 AWWNNQWVFTTWREDTRPGDQFQHTKKFCFDIAISHTSPVTLJDCISWKGKQWIK-RDKT 554  
 Db 522 --GNQVFATVTKRQQIMSDD----LCUDASSNGPWNWRCMGGNQEWVYDAEKW 572  
 Qy 555 LYHPSVSGSMCSDSDFHRIFTN--CNPSII--TQWLFB 590  
 Db 573 IRHTNSGQCLQRATRD--ANTPLURPCSYGKQGQWLM 609

GL13 MOUSE STANDARD, PRT; 556 AA.  
 ID GL13\_MOUSE\_STANDARD; AC Q8C93; Q8BL4; Q8Y73;  
 AC Q8C93; Q8BL4; Q8Y73;  
 DT 01-OCT-2004 (Rel. 45, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Polypeptide (Protein-UDP acetylgalactosaminyltransferase 13) (UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase 13) (Polypeptide GalNAc:polypeptide N-acetylgalactosaminyltransferase 13) (Polypeptide GalNAc transferase 13) (GalNAc-T13) (pp-Gantase 13).  
 Name=GalN13;  
 OS Mus musculus (Mouse).  
 OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. OX NCBITaxonID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND TISSUE SPECIFICITY.  
 RX TISSUE=Brain; PubMed=1240714; DOI=10.1074/jbc.M203034200;  
 RA Zhang Y., Iwasaki H., Wang H., Kudo T., Kalka T.B., Hennet T., Kubota T., Cheng L., Inaba N., Gotoh M., Togayachi A., Guo J.-M., Hidatomi H., Nakajima S., Nakamura M., Marth J.D., Narimatsu H.;  
 RA "Cloning and characterization of a new human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase, designated pp-GalNAc-T13, that is specifically expressed in neurons and synthesizes GalNAc alpha-serine/threonine antigen.", J. Biol. Chem. 278:573-584 (2003).  
 [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC STRAIN=CS7BL/6J; TISSUE=Corpora quadrigemina, and Hypothalamus;  
 RX MEDLINE=22354683; PubMed=14466851; DOI=10.1038/nature0166;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Nikaido I., Osato N., Saito R., Suzuki H., Yamamoto T., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T., Baldarelli R., Hill D.P., Bolt C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batyalov S., Bilezikian K.W., Blake J.A., Bradt D., Brusic V., Chochia C., Corbani L.B., Cousins S., Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S., Gasterbertrand T., Garibaldi M., Glassi C., Godzik A., Gough J., Grimmond S., Gustincich S., Hirokawa N., Jackson J.J., Jarvis B.D., Kanai A., Kawaji Y., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Nutata K., Okuda T., Pavan W.J., Peretea G., Pesole G., Petrovsky N., Pilai R., Pontius J.U., Qi D., Ramachandran S., Ravid S., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K., Sultana R., Takeuchi Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wrayshaw Boris A., Yangisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hizozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraishi T., Waki K., Kawai J., Arakawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imoto K., Iishi Y., Itoh M., Kagaawa I., Miyazaki A., Sasaki K., Sasada D., Shibata K., Shinagawa A., Yesunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayasizaki Y.;  
 RA "Analysis of the mouse transcriptome based on functional annotation of full-length cDNAs.", Nature 420:563-573 (2002).  
 RL [3]  
 RP FUNCTION  
 RX PubMed=811846;  
 RA Hennet T., Hagen F.K., Tabak L.A., Marth J.D.;  
 RT "T-cell specific deletion of a polypeptide N-acetylgalactosaminyltransferase gene by site-directed recombination.", Proc. Natl. Acad. Sci. U.S.A. 92:12070-12074 (1995).  
 RN [4]  
 RP TISSUE SPECIFICITY.  
 RX Published=1265184; DOI=10.1093/glycob/cwg062;  
 RA Young W.W. Jr., Holcomb D.R., Ten Hagen K.G., Tabak L.A.;  
 RT "Expression of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase isoforms in murine tissues determined

RT by real-time PCR: a new view of a large family";  
 RL Glycobiology 13:549-557 (2003).  
 CC -!- FUNCTION: Catalyzes the initial reaction in O-linked  
 oligosaccharide biosynthesis, the transfer of an N-Acetyl-D-  
 galactosamine residue to a serine or threonine residue on the  
 protein receptor. Has a much stronger activity than GalNT I to  
 transfer GalNAc to mucin peptides, such as Muc5AC and Muc7. Able  
 to glycosylate SDC3. Probably responsible for the synthesis of Tn  
 antigen in neuronal cells.  
 CC -!- CATALYTIC ACTIVITY: UDP-N-Acetyl-D-galactosamine + polypeptide =  
 UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -!- COFACTOR: Manganese and calcium (By Similarity).  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
 Similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 Event=Alternative splicing; Named isoforms=2;  
 Name=1; Isoform=08CF93-1; Sequence=Displayed;  
 Name=2; Isoform=08CF93-2; Sequence=ISP\_011220;  
 Note=No experimental confirmation available;  
 CC -!- TISSUE SPECIFICITY: Specifically expressed in neuronal cells. Not  
 expressed in glial cells such as astrocytes. Expressed at low  
 level.  
 CC -!- DOMAIN: There are two conserved domains in the glycosyltransferase  
 region: the N-terminal domain (domain A, also called Gal1 motif),  
 which is probably involved in manganese coordination and substrate  
 binding and the C-terminal domain (domain B, also called  
 Gal/GalNAc-T motif), which is probably involved in catalytic  
 reaction and UDP-Gal binding (By Similarity).  
 CC -!- DOMAIN: The ricin B-type lectin domain binds to GalNAc and  
 contributes to the glycoprotein specificity (By Similarity).  
 CC -!- MISCELLANEOUS: The absence of GalNT3 (which was wrongly assigned  
 as GalNT8), does not induces clear defects in mice. It however  
 abolishes Tn antigen in neuronal cells.  
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
 subfamily.  
 CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 in the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).

CC EMBL: AB082928; BAC54546.1; -.  
 DR EMBL; AK038387; BAC29981.1; -.  
 DR EMBL; AK05417; BAC253.1; -.  
 DR MGI; MGI:2139447; Galnt13.  
 DR InterPro; IPR001173; Glyco\_trans\_2.  
 DR InterPro; IPR008997; Ricin\_B\_lectin.  
 DR InterPro; IPR000772; Ricin\_B\_lectin.  
 DR Pfam; PF00535; Glycos\_transf\_2\_1.  
 DR Pfam; PF00652; Ricin\_B\_lectin\_3.  
 DR SMART; SM00458; RICIN\_1.  
 DR PROSITE; PS00231; RICIN\_B\_LRCTIN\_1.  
 KW Alternative splicing; Calcium; Glycosyltransferase; Golgi stack;  
 Lectin; Manganese; Signal-anchor; Transferase; Transmembrane.  
 PT DOMAIN 1 4 Cyttoplasmic (Potential); Transmembrane.  
 PT TRANSMEM 5 27 Signal-anchor for type II membrane protein (Potential).  
 PT DOMAIN 28 556 Luminal (Potential).  
 PT DOMAIN 114 224 Catalytic subdomain A.  
 PT DOMAIN 284 346 Catalytic subdomain B.  
 PT DOMAIN 428 550 Ricin B-type lectin.  
 PT DISULFID 441 458 By similarity.  
 PT DISULFID 481 496 By similarity.  
 PT DISULFID 522 539 By similarity.  
 PT CARBOHYD 94 104 N-linked (GlcNAc. . .) (Potential).  
 PT CARBOHYD 116 116 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 551 551 N-linked (GlcNAc. . .) (Potential).  
 FT VARSPLTC 466 466 V -> VNPICLSPSUGVAGBCCSNHFLYGLVPTINEQ  
 FT V (in isoform 2).  
 FT /FTid=ISP\_011220.  
 FT CONFLICT 457 457 N -> K (in Ref. 2; BAC32353).  
 FT SEQUENCE 556 AA: 63982 MW: 59394CFD0AED148 CR054;  
 SQ 88 GNGEQGRP--YPMTDAERVDQAYRNGENPVISDKISNRSLPDIRHNCNSKRYLETP 145  
 QY 56 GPGEMWKVLUPLKDOEHNKELPKINGPNLMSDLIANSPLDVRLGCKTKVYPDLP 115  
 Db 146 NTSTIIPPHNEGSSLRATVHSVLRNRSRPELAVBLVUDDFSDREHKKPLEDY-MALPP 204  
 Db 116 NTSVVIVHNEAWNSTLRTVHSVLRNPHLSEVINDASERDPLKLTENYKVE 175  
 Db 205 SVHLTRKREGIIRTRMAGSATGVITFLSDHCRANVNLPLPFLDRIARNRKTIVCP 264  
 Db 176 PVKLRIMBERSGLRARIKGAAASKGQVTFDADHCECTLGWLEPLLARIKEDRKTIVCP 235  
 QY 255 MIDVHDHDFRYTQAG-DAMGARDHMYXKIRPP--ELOKABPSDPRSPVAGG 320  
 Db 236 IIBVVISDTFPEY--MACSDMTYQGENWNLNFNMRVPPVPREMDRKGDRITLPVPTMAGG 293  
 QY 321 LFAVDRKFWELGGYDPEGLINWQGQEVILSKVUMCGARMEDIPCSRQVGHIVKVPKV 380  
 Db 294 LFSIDRNRFLFEGIYDQGDINGENIEMSFRIWQCCSLEVTCSHVGVRKATPTP 353  
 QY 381 PAGVS--IARNLGRVAETWMDAYEAYTIVRORRBYRHLSAGDVAVOKURSSLNCKSFNPW 438  
 Db 354 PGGTGHVIVNKNNRKLAEWMDKFDFFVITISPCVVKVQYDGVSVRKTRLENLCKPPSWY 413  
 QY 439 MTKIAANDPKFVPPVRAAM--GEFRNGTGLCAPIKKGALGSPLLEGURGRBAA 496  
 Db 414 LENI-----YEDSOFRRYVSLGIRNVTNOCNDNGKRENKEVKGIFNC--HGMG-- 462  
 QY 497 WNNQMVFTFWREDIRPDPQHPTKKFCDAISHTSPVLYDCISMKGQLWQKVKDK-TL 555  
 Db 453 --GNOVFSYTADEKIRD---LQDVSLSGPVIMLKHMRCNQWYDAEKTIL 514  
 QY 556 YHPVSGND-CSESDEH-I-FMTCNCPSLSDWLFER 591  
 Db 515 RHANSNOQCLDEPSEEDKVWPTMDCS-GSRSQOQLRN 551

Search completed: November 22, 2004, 13:43:40  
 Job time : 71 secs

This Page Blank (uspto)